



UNIVERSIDAD DE JAÉN
**FACULTAD DE CIENCIAS
EXPERIMENTALES
DEPARTAMENTO DE ESTADÍSTICA
E INVESTIGACIÓN OPERATIVA**

TESIS DOCTORAL
**ANÁLISIS DE REGRESIÓN CON DATOS
IMPRECISOS: UN NUEVO ENFOQUE QUE
UTILIZA DISTANCIAS DIFUSAS Y SUS
APLICACIONES**

**PRESENTADA POR:
MARÍA DE LA CONCEPCIÓN AGUILAR PEÑA**

**DIRIGIDA POR:
DRA. DÑA. CONCEPCIÓN BEATRIZ ROLDÁN LÓPEZ DE
HIERRO
DR. D. ANTONIO FCO. ROLDÁN LÓPEZ DE HIERRO**

JAÉN, 30 DE JULIO DE 2015

ISBN 978-84-8439-935-3

Universidad de Jaén

Departamento de Estadística e Investigación Operativa



**Análisis de regresión con datos imprecisos:
Un nuevo enfoque que utiliza distancias difusas
y sus aplicaciones**

Memoria de TESIS DOCTORAL realizada bajo la dirección de los profesores doctores Concepción Beatriz Roldán López de Hierro y Antonio Francisco Roldán López de Hierro y presentada por la Doctoranda María de la Concepción Aguilar Peña para optar al grado de Doctora en Estadística:

Fdo: María de la Concepción Aguilar Peña

VºBº de los directores:

Fdo: Dra. D^a. Concepción B. Roldán y Fdo: Dr. D. Antonio F. Roldán

La presente Tesis Doctoral está avalada a la fecha de su lectura por los siguientes artículos publicados en revistas del JCR:

C. Aguilar-Peña , A.F. Roldán-López-de-Hierro, C. Roldán-López-de-Hierro, J. Martínez-Moreno, Estimation of a fuzzy regression model using fuzzy distances, *IEEE Transactions on Fuzzy Systems*. Factor de impacto: 8.746 (Posición 1 de 123 en la categoría 'Computer Sciences, Artificial Intelligence', y posición 1 de 249 en la categoría 'Engineering, Electrical & Electronic' según JCR 2014).

C. Aguilar-Peña, A.F. Roldán-López-de-Hierro, C. Roldán-López-de-Hierro, J. Martínez-Moreno. A family of fuzzy distance measures of fuzzy numbers. *Soft Computing*. D.O.I.: 10.1007/s00500-014-1497-0. Factor de impacto: 1.271 (Posición 58 de 102 en la categoría 'Computer Science, Interdisciplinary Applications' según JCR 2014, actual).

Agradecimientos

Dedicado a mi marido e hijos. Gracias por vuestro apoyo y cariño.

Quiero también dar las gracias a mis supervisores, Dra. D^a Concepción B. Roldán López de Hierro y Dr. D. Antonio F. Roldán López de Hierro por su paciencia y dirección en esta memoria.

Índice

Resumen	v
Abstract	vii
1. Introducción	1
2. Preliminares	13
2.1. El modelo de regresión clásico	13
2.1.1. Planteamiento del problema de regresión	14
2.1.2. Las funciones de pérdida	15
2.1.3. El método de mínimos cuadrados para la regresión lineal	17
2.1.4. Medidas de bondad del ajuste	20
2.1.5. Diagnóstico en el modelo de regresión lineal	22
2.1.6. Selección de variables explicativas	24
2.2. Teoría de Conjuntos Difusos	26
2.2.1. Valores Difusos	26
2.2.2. Algunas clases de números difusos	30
2.2.3. Operaciones entre números difusos	35

2.2.4. Variables aleatorias difusas	38
2.2.5. Valor esperado de una variable aleatoria difusa	39
3. Medida de la distancia entre números difusos	41
3.1. Medidas de distancias entre conjuntos arbitrarios	42
3.2. Medidas de distancias entre intervalos compactos	48
3.3. Medidas de distancia entre números difusos	53
3.3.1. Un orden parcial sobre los números difusos trapezoidales	53
3.3.2. Medidas de similitud entre números difusos	55
3.4. Aplicaciones	71
3.4.1. La topología inducida sobre \mathcal{T}	71
3.4.2. Estudio comparativo con los resultados de otros autores .	74
3.4.3. Ranking	76
4. Métodos de regresión difusa	79
4.1. Un primer acercamiento al problema de regresión difusa	79
4.2. El planteamiento del problema de regresión difusa	84
4.3. Una nueva metodología de regresión difusa	90
4.3.1. El procedimiento de estimación cuando las amplitudes de- penden de las variables explicativas	92
4.3.2. El procedimiento de estimación cuando alguno de las am- plitudes no depende de las variables explicativas	93
4.3.3. Una extensión de la metodología previa	99
4.4. Medidas de bondad de ajuste	99

5. Ejemplos ilustrativos	103
5.1. Resumen de la metodología propuesta	103
5.2. Ejemplo 1: Amplitudes dependientes de variables explicativas . .	106
5.3. Ejemplo 2: Datos ampliamente utilizados	114
5.4. Ejemplo 3: Amplitudes independientes de variables explicativas .	120
5.5. Ejemplo 4: Las variables de entrada son difusas	127
5.6. Ejemplo 5: 100 % de variabilidad difusa explicada	130
6. Conclusiones y observaciones finales	133
Bibliografía	137

Resumen

El análisis de regresión es una herramienta estadística potente con muchas aplicaciones en diferentes áreas. Este problema en un entorno difuso ha sido tratado en la literatura desde diferentes puntos de vista y teniendo en cuenta una variedad de datos de entrada/salida (reales o difusos). En esta memoria se presenta una nueva metodología basada en una familia de medidas de distancia difusas entre números difusos arbitrarios que se define utilizando algunas de las características posibilistas y geométricas más importantes de cualquier número difuso. A continuación, en este contexto y utilizando el método de mínimos cuadrados se propone una nueva técnica de regresión difusa para resolver problemas lineales y no lineales. Este proceso de estimación, en general, se puede considerar fácil de aplicar en la práctica y no se limita a números difusos triangulares. Finalmente, algunos ejemplos numéricos ilustran su utilidad y aplicabilidad.

Abstract

Regression analysis is a powerful statistical tool which has many applications in different areas. This problem under a fuzzy environment has been treated in the literature from different points of view and considering a variety of input/output data (crisp or fuzzy). In this study we present a new methodology based on a family of fuzzy distance measures between arbitrary fuzzy numbers which involve some of the most important possibilistic and geometric characteristics of any fuzzy number. Next in this context and using the method of least squares we propose a new fuzzy regression technique to solve linear and non-linear problems. This estimation process, in general, can be considered easy to apply in practical situations and it is not limited to triangular fuzzy numbers. Finally, numerical examples are provided to illustrate its usefulness and applicability.

CAPÍTULO 1

Introducción

El principal objetivo de la presente Memoria es el de introducir un nuevo método de regresión difusa basado en criterios de decisión difusos y cuyos resultados sean tan buenos, e incluso mejores, que los obtenidos empleando procedimientos reales. Como veremos, esta nueva perspectiva nos obliga a afrontar diversos problemas que, habiendo sido ya tratados inconexamente en algunos casos en la literatura existente en el tema, nunca se han profundizado lo suficiente como para dar una respuesta satisfactoria a este problema desde un punto de vista genuinamente difuso.

Una de las principales tareas de todas las ciencias naturales y sociales, en general, es la de encontrar las posibles relaciones que pueden existir entre las magnitudes que intervienen en los procesos físicos que son objeto de estudio de esa disciplina. Las técnicas actuales proporcionan una amplia gama de posibilidades para abarcar fenómenos cada vez más complejos, donde intervienen cada vez más factores explicativos y magnitudes que, en ocasiones, pueden resultar

difíciles de medir (por ejemplo, a nivel subatómico).

El análisis de regresión es una potente herramienta que tiene una amplia variedad de aplicaciones en muchas áreas de investigación. Esta metodología se centra en encontrar las posibles relaciones que puedan existir entre una variable dependiente y una o más variables independientes. Los modelos así obtenidos son ampliamente utilizados en procesos de predicción y previsión.

Los primeros estudios sobre regresión datan del siglo XIX, cuando Sir Francis Galton afrontó el problema de construir un modelo afín para explicar la relación entre dos variables reales cuantitativas. Con el paso del tiempo, el análisis de regresión ha adquirido una enorme importancia dentro de las ciencias en general, que, normalmente, ya no pueden ser concebidas sin tener en cuenta la necesidad de realizar estudios empíricos para la búsqueda de modelos que nos lleven a comprender qué variables independientes son las que mejor explican el comportamiento de la variable dependiente.

Los fundamentación básica sobre la que se construye el análisis de regresión (al menos, su vertiente paramétrica) es la siguiente: cuando se pretende estudiar la influencia de una o más variables independientes (o *explicativas*) sobre el comportamiento de otra variable dependiente (o *explicada*), hay que partir de diversas consideraciones. Éstas son, a menudo, geométricas y físicas, y nos llevan a plantearnos la posibilidad de que un modelo concreto, dependiente de ciertos parámetros, pueda explicar, en mayor o menor medida, los hechos que acontecen en el desarrollo del proceso natural. A continuación debe desarrollarse una observación lo más detallada posible del mencionado proceso, tomando muestras aleatorias de las magnitudes que nos parezca que pueden influir en el mismo. Teniendo en cuenta una medida de la diferencia que puede existir entre los datos reales que se han observado empíricamente y los que pudieran postularse a nivel teórico, en muchas ocasiones es posible estimar aquellos parámetros del modelo que mejor explican el proceso, en el sentido de que el error global que se cometa sea el mínimo posible.

Si nos detenemos a analizar el procedimiento que acabamos de comentar, cabe hacer las siguientes reflexiones.

- En primer lugar, es importante evitar hacer difuso lo que no lo es o determinar como cierto o aleatorio aquello que es difuso. Por ello, conviene realizar una adecuada distinción entre incertidumbre y aleatoriedad.
- La necesidad de refinar cada vez más los experimentos desarrollados con objeto de obtener modelos más precisos, nos lleva que la toma de datos no puede ser observada como un detalle menor. Las magnitudes que se emplean en la ciencias son cada vez más complejas y demandan de instrumentos de medida y de simulación cada vez más sofisticados. Por ejemplo, cabe citar la exploración aeroespacial, la astronomía y la mecánica cuántica como ejemplos de campos de estudio en los que la precisión de los aparatos de medida juega un papel cada vez más importante si queremos obtener resultados relevantes. Superar esta dificultad, en ocasiones, lleva a trabajar con variables que ya no toman un valor real concreto. De esta forma, partimos de la premisa fundamental de que los valores exactos de ciertas magnitudes son, intrínsecamente, inciertos y, por ello, lo más razonable es considerar que dichos valores están comprendidos dentro de ciertos límites razonables. Surge así la necesidad de obtener conclusiones basadas en variables propiamente imprecisas y de emplear *datos difusos*.
- La aleatoriedad o incertidumbre estocástica del resultado experimental es otra fuente de incertidumbre a tener en cuenta. Gran cantidad de experimentos no producen los mismos resultados cuando son sometidos a condiciones iniciales semejantes. Existe un grado de aleatoriedad en la experimentación que entraña la necesidad de considerar modelos probabilísticos, cuyos resultados no son valores reales concretos, sino que existe un margen de error propiamente intrínseco al experimento, y que no se explica necesariamente por la imprecisión de los aparatos de medida. Este

tipo de incertidumbre se formaliza a través de de la Teoría de la Probabilidad y según Zadeh [70], esta teoría por sí sola no aborda el concepto de parcialidad y no cuenta con los mecanismos necesarios para presentar el significado de percepciones y realizar cálculos con dichas representaciones. En teoría de la probabilidad, un proceso es estocástico o no estocástico; una serie temporal es estacionaria o no estacionaria, un evento sucede o no sucede; los eventos A y B son independientes o no independientes; y así sucesivamente.

- Teniendo en cuenta lo anterior, en muchos problemas de la vida real, el analista se encuentra la presencia simultánea de datos ciertos (se conocen todos sus posibles valores, numéricos o no numéricos que afectan a cada estado), aleatorios (sus posibles valores se ven afectados por leyes de probabilidad) e inciertos (no pueden obtenerse leyes de probabilidad debido al conocimiento parcial o al desconocimiento de dicha ley). Es por ello, que hay situaciones en las que contamos con datos medibles y otras veces sólo disponemos de datos valiables.
- No es sencillo manejar esta *incertidumbre* (también denominada *imprecisión, vaguedad o ambigüedad* en muchos contextos). Una de las perspectivas más adecuadas para hacerlo es la *teoría de conjuntos difusos* (sobre cuyos fundamentos volveremos más adelante a lo largo de la presente Memoria), que ofrece la posibilidad de generalizar los números reales a *números difusos*. Éstos son entidades probabilísticas que determinan el grado de certidumbre que se puede tener de que cierto valor de una variable tome valores dentro de un intervalo concreto.
- En este contexto, la gestión de la incertidumbre cuenta con este nuevo enfoque que no debe de entenderse como competitivo respecto al enfoque probabilístico sino complementario. Lo incierto y lo aleatorio se unen sin inconveniente ninguno e incluso se puede afirmar que ambos convergen a

lo que Zadeh [70] denomina *razonamiento probabilístico basado en la percepción* dando así lugar a conceptos como el de variable aleatorias difusa, proceso estocástico difuso, etc.

- Dar un enfoque difuso a un problema planteado evita la pérdida de información que se produce con el uso de métodos estadísticos tradicionales que trabajan con valores promedios ocultando los valores extremos y permite entender la realidad incorporando información de gran valor, tales como juicios razonados acerca de lo que va a suceder. En este sentido, los modelos de regresión difusa son muy útiles para establecer rangos de incertidumbre.
- Uno de los principales ingredientes en teoría de regresión es la noción de *medida de la distancia*. Cuando se menciona la palabra “distancia”, nos viene a la cabeza irremediamente la distancia euclídea, que es la que utilizamos todos los días en el mundo real. Aplicada a números reales, la distancia que los separa es la longitud del intervalo cuyos extremos son esos números reales, es decir, la diferencia, en valor absoluto, entre los dos números. Cuando tratamos de trasladar esta noción al campo difuso, no encontramos una respuesta plenamente satisfactoria. Aunque Mizumoto y Tanaka [40] establecieron en su momento una aritmética básica entre números difusos que extiende a la aritmética real, no es posible determinar una noción de *diferencia de números difusos* que sea coherente con la incertidumbre de los procesos que tratan de describir. En concreto, no se ha determinado ninguna forma de restar números difusos que sea compatible con la suma de los mismos, es decir, si a una diferencia de números difusos le sumamos el sustraendo, nunca obtenemos el minuendo. De esta forma, las operaciones elementales entre números difusos establecen, más bien, una estructura semilineal, donde las principales operaciones a considerar o, al menos, las que más sentido empírico tienen, son la suma y el producto por escalares no negativos.

- La ausencia de una definición básica de diferencia de números difusos, coherente con la suma de los mismos, provoca inconsistencia al tratar de definir la distancia entre dos números difusos como la diferencia entre ellos. De hecho, no es nada sencillo tratar de encontrar una métrica, con aplicaciones prácticas, definida sobre todo el conjunto de números difusos, ni siquiera tomando valores reales (de todas formas, esto sería un poco absurdo, pues estaríamos tratando de dar un valor exacto para determinar la distancia entre dos cantidades inciertas). En la literatura existente en el tema pueden encontrarse muchos artículos en los que los autores anuncian la intervención de métricas reales pero, a la hora de la verdad, éstas no son métricas en sentido estricto: por ejemplo, a menudo es posible encontrar dos números difusos diferentes que, aplicando la metodología propuesta, están a distancia cero el uno del otro.
- Las más modernas aproximaciones a este problema no mencionan explícitamente la palabra *métrica*, sino que más bien hablan de *medida de similitud* (véase, por ejemplo, [61]). Esta nomenclatura responde con mayor claridad al procedimiento desarrollado pues, en cierta forma, tratan de dar una medida de la semejanza o de la diferencia entre dos números difusos, pero no tratan de describir un proceso que garantice cumplir los cuatro axiomas tradicionales que se emplean para definir la noción de distancia. Estas medidas de similitud suelen basarse en consideraciones geométricas y/o analíticas dependiendo de las formas de los números difusos considerados. No obstante, en muchas ocasiones, suelen proporcionar números reales como medidas de similitud lo cual, nuevamente, es incoherente con la visión probabilística del fenómeno estudiado.
- Por último, hemos de comentar que cualquier procedimiento que trate de buscar los parámetros de un modelo que, globalmente, produzcan el *menor* error posible deben, al menos, definir en qué sentido utilizan la palabra “menor”. Desafortunadamente, aunque se han introducido muchos órdenes

parciales en el conjunto formado por todos los números difusos, ninguno de ellos ha ofrecido resultados plenamente satisfactorios en todos los campos de estudio en los que se ha empleado. De hecho, si existiese un orden parcial comúnmente aceptado, éste jugaría un papel esencial en ámbitos tan importantes como la *toma de decisiones*, área en la que se han hecho grandes avances en los últimos años, especialmente en el campo difuso. Ni que decir tiene que la necesidad de un orden parcial también se pone de manifiesto cuando tratamos de buscar una noción de métrica entre números difusos que verifique la archiconocida *desigualdad triangular*.

Uno de los primeros resultados que explícitamente condujeron a la necesidad de introducir una nueva visión probabilística de algunos fenómenos naturales fue el *principio de incertidumbre de Heisenberg* (1927), que establece la imposibilidad física de conocer, a la vez y con toda exactitud, la posición y la velocidad de una partícula elemental (por ejemplo, de un electrón). A grandes rasgos, podemos explicar este principio de la siguiente forma: si se desea conocer la posición exacta de una partícula elemental, podemos enviarle un haz de fotones. Al chocar un fotón con el electrón, conoceremos la posición exacta, en ese momento, del electrón. Pero el choque modificará el vector velocidad del electrón de una forma incontrolable de tal manera que, con total precisión, ya no conoceremos su velocidad después del choque.

Esta incertidumbre nos lleva a plantearnos si un número real es la mejor forma de medir magnitudes físicas, o más bien éstas han de ser interpretadas desde un punto de vista probabilístico. Uno de los primeros en generalizar el concepto de distancia real haciendo intervenir un sentido probabilístico fue K. Menger. En 1942, Menger introdujo la noción de *espacio métrico estadístico*. Años después de su primer intento, modificó sus axiomas de una forma más sencilla de interpretar: la distancia entre dos puntos del espacio dejó de ser un número real, y pasó a ser una *función de distribución probabilística* que mide, para cada número real x , la probabilidad de que la distancia entre dos puntos

dados sea menor que x . Él llamó a estas nuevas estructuras *emsemble flous*, que se traduce del francés como *conjuntos difusos* o *borrosos*.

Pero el verdadero fundador de la teoría de los conjuntos difusos es, indudablemente, L.A. Zadeh. Esta teoría está basada en la idea de que no siempre es posible determinar con total exactitud si un elemento pertenece o no a un conjunto. Matemáticamente, la idea de Zadeh [70] puede ser formalizada como sigue. Ya era conocido con anterioridad que la pertenencia de un elemento a un conjunto puede ser equivalentemente expresada utilizando una cierta aplicación llamada *función característica del conjunto*. Cada subconjunto A de un cierto universo \mathcal{U} viene completamente determinado por la aplicación $f_A : \mathcal{U} \rightarrow \{0, 1\}$ que vale 1 sobre los elementos de A y 0 sobre el resto de elementos. Si se reemplaza el conjunto discreto $\{0, 1\}$ por el intervalo $\mathbb{I} = [0, 1]$, entonces el valor de $f_A(x)$ puede expresarse como la *probabilidad* (o el *grado de certeza* que se tiene) de que el elemento x pertenezca al conjunto A . Así, un *conjunto difuso* no es más que una función $F : \mathcal{U} \rightarrow [0, 1]$ que expresa, de cierta forma, la *incertidumbre* que tenemos de que un elemento pertenezca, o no, a un conjunto. A partir de la idea pionera de Zadeh se han propuesto cientos de trabajos en los que la naturaleza imprecisa en el comportamiento de un sistema dado responde mejor a un modelo difuso que a un modelo real estocástico. Y hoy posiblemente sean millones los trabajos que se han desarrollado en el campo de la teoría de los conjuntos difusos y de sus aplicaciones. En la actualidad, este campo de investigación posee una gran relevancia tanto por sus implicaciones matemáticas o teóricas como por sus aplicaciones prácticas. Esta teoría se ha utilizado con éxito para tratar la información imprecisa, incompleta o imprecisa, así como la incorporación de un juicio subjetivo de expertos en problemas de análisis.

Utilizando la idea de conjunto difuso, se han propuesto, a lo largo del tiempo, multitud de posibles definiciones acerca del concepto de *espacio métrico difuso* (cabe citar, por ejemplo, las aproximaciones dadas por I. Kramosil y J. Michálek [34], A. George y P.V. Veeramani [23], F. Castro-Company y S. Romaguera [7],

O. Kaleva y S. Seikkala [32], Roldán *et al.* [51], y, en un ambiente más probabilístico, por B. Schweizer y A. Sklar [56]). Sin embargo, ninguna concepción trata, de manera particular, el problema de cómo definir una noción de distancia entre dos números difusos arbitrarios. Desde nuestro punto de vista, con objeto de no perder información y de respetar el carácter impreciso de los datos que se manejan, no cabe duda de que la distancia entre dos números difusos debe de ser otro número difuso. Para nosotros, ésta es una condición indispensable. Sin embargo, no está claro cómo considerar una forma canónica para asociar un número difuso a dos números difusos que se pueden interpretar como la cercanía o lejanía entre ellos. De hecho, no están claras las propiedades que debemos exigir a una aplicación como lo que buscamos. El problema es doble: por un lado, siguiendo la noción clásica, una métrica en un conjunto debe tomar valores en el intervalo de los números reales no negativos, pero no está clara la noción de “no negatividad” entre números difusos; por otro lado y relacionado directamente con lo anterior, al carecer de un orden parcial que pueda decidir acerca de la “positividad”, o no, de un número difuso, la desigualdad triangular se transforma en un axioma difícilmente generalizable.

Sin embargo, teniendo en cuentas sus posibles aplicaciones (véase, por ejemplo, [8, 22, 46, 62, 63, 64]), las cuales reflejan la importancia de encontrar una buena definición, es habitual relajar algunas de estas condiciones exigibles para conseguir formas más generales para determinar la semejanza/diferencia entre dos números difusos. Cabe entonces hablar de *medidas de similitud*, donde los axiomas que se exigen, que parten más bien de consideraciones geométricas, son menos estrictos. Así, se puede hablar de *semimétricas*, *pseudométricas*, *ultramétricas*, *quasimétricas*, etc.

Si queremos introducir una metodología para evaluar la semejanza entre dos números difusos, lo que parece razonable es interpretar tal valor como un número difuso. Muchos investigadores han trabajado con medidas de distancia entre números (véase, por ejemplo, [3, 4, 13]), pero muy pocos de ellos han utiliza-

do esta interpretación dentro del conjunto de todos los números difusos. Por ejemplo, una medida (muy general) de la distancia trapezoidal difusa entre dos números difusos se desarrolla en [27]. Pero, en todos los casos, la distancia sólo se define para formas simples de números difusos, tales como números difusos trapezoidales, ya que el uso de formas menos regulares puede llevarnos a cálculos muy complicados.

Uno de los objetivos principales de esta Memoria, y al que dedicamos íntegro el Capítulo §3, es el de presentar una familia de medidas de similitud en el conjunto formado por todos los números difusos. Esta familia no se restringe únicamente a números difusos trapezoidales, sino que abarca el conjunto completo de números difusos. Por lo tanto, ofrecemos muchas formas diferentes para obtener un número difuso que se pueda interpretar como la cercanía entre dos números difusos. Nuestra metodología está basada en hacer intervenir algunos de los elementos posibilistas y geométricos más importantes de cualquier número difuso para definir un nuevo número difuso. Las principales ventajas de los elementos de la familia de medidas de similitud que vamos a introducir son las siguientes:

- se definen en el conjunto de todos los números difusos (no sólo entre números difusos trapezoidales);
- algunos de los subconjuntos más útiles del conjunto de todos los números difusos (incluyendo números difusos triangulares y trapezoidales) son cerradas bajo estas medidas de similitud;
- en algunos casos, como en conjunto de los números difusos trapezoidales, son auténticas métricas difusas;
- en estos casos, este conjunto está dotado de una topología Hausdorff que verifica el primer axioma de numerabilidad;

- por otra parte, una vez que hemos fijado los elementos geométricos que intervienen en nuestra definición, el cálculo de la medida de la distancia es un proceso muy fácil, intuitivo y con un coste computacional muy bajo;
- finalmente, esta familia nos permite resolver problemas que se plantean en un ambiente difuso aportando nuevas soluciones, en unos casos, y en otros obteniendo resultados tan buenos o similares a los obtenidos por otras técnicas más complejas.

En relación a éste último punto, la familia de medidas de similitud que introducimos nos permite afrontar, de forma novedosa, el problema de regresión difusa en el siguiente sentido. Partimos del hecho de que, en general, la mayoría de los trabajos de investigación que se han desarrollado en este campo presentan un conflicto entre el uso de medidas de distancia reales y la interpretación de los datos difusos como distribuciones de posibilidad. En este sentido, la solución que proponemos en esta Memoria, considerando una medida difusa de similitud, es compatible con esta interpretación. Además el proceso asociado de estimación numérica es muy fácil de aplicar en comparación con otros métodos que son relativamente complejos. La principal contribución de la nueva formulación que presentamos en esta Memoria es ser capaz de considerar distancias difusas, de obtener parámetros difusos fácilmente para desarrollar modelos de regresión lineal y no lineal, y considerar una amplia gama de datos difusos.

Esta Memoria está organizada de la siguiente manera:

- En el primer capítulo se presentan algunas nociones básicas de la teoría de regresión y de la teoría de conjuntos difusos.
- En el segundo capítulo se introduce la primera parte esencial de nuestra investigación: como alternativa a la noción de métrica real, se proponen nuevos conceptos de “métricas” sobre un conjunto arbitrario, y que deben

tomar valores en el mismo conjunto, y se estudian sus principales propiedades. Este capítulo se completa con un estudio comparativo que permite ver cómo actúa el orden parcial introducido en relación a otros órdenes parciales considerados con anterioridad por otros autores.

- En el tercer capítulo, en base a los resultados del capítulo anterior, se introduce la segunda parte esencial de nuestro trabajo de investigación, que puede ser considerado como el objetivo principal de la presente Memoria: proponer una nueva técnica de regresión difusa siguiendo el método de mínimos cuadrados y utilizando las medidas de distancia difusas introducidas en el capítulo anterior, que permita dar al problema planteado un enfoque flexible, íntegramente difuso y que sea fácil de entender por investigadores de otras áreas.
- En el cuarto capítulo se ilustra la aplicación del análisis de regresión difuso a datos reales y se realiza un estudio comparativo con los resultados obtenidos por otros autores.
- Finalmente, en el quinto capítulo, se incluyen las conclusiones principales que hemos obtenido con nuestro trabajo de investigación, así como algunas posibles líneas de investigación futuras, pendientes de afrontar.

CAPÍTULO 2

Preliminares

En este capítulo vamos a presentar las nociones básicas y los principales resultados que nos harán falta a lo largo de esta memoria. En primer lugar, haremos un repaso a las principales técnicas empleadas en el Análisis de Regresión Clásico y, a continuación, describiremos algunas de las principales características de la Teoría de Conjuntos Difusos.

2.1. El modelo de regresión clásico

El análisis de regresión es una técnica estadística que se emplea para estudiar las relaciones que pueden existir entre diferentes variables aleatorias, analizando y modelando dichas relaciones a través de numerosos métodos.

Este método comenzó a desarrollarse con los estudios realizados por Sir Francis Galton (1822-1911) acerca del análisis de regresión lineal. Dicha técnica juega un papel muy importante en la teoría de la Estadística ya que proporciona un procedimiento básico para desarrollar una gran variedad de técnicas estadísticas esenciales como el análisis discriminante, el diseño de experimentos, etc.

Desde un punto de vista general, dado que no se conoce la verdadera forma

del proceso de generación de datos, el análisis de regresión depende, en gran medida, de las suposiciones que hagamos acerca de cómo se van obteniendo los datos. Por ello, en las últimas décadas, se han desarrollado numerosos métodos de regresión para dar respuesta a las diferentes necesidades que van surgiendo en la práctica. Para facilitar este trabajo, existen numerosos programas estadísticos que han contribuido, en gran medida, a la comprensión del proceso desarrollado y al uso de esta técnica por investigadores/as que no son expertos/as en este tema. Los modelos obtenidos permiten plantear desde soluciones muy simples hasta modelos bastante más complejos y sofisticados.

En esta sección presentamos, de una forma simple, el problema general del análisis de regresión. En lo que sigue, denotaremos por \mathbb{R} al conjunto formado por todos los números reales, y \mathbb{R}^N denotará el producto cartesiano de N copias idénticas de \mathbb{R} .

2.1.1. Planteamiento del problema de regresión

En general, el problema de regresión consiste en encontrar las relaciones que pueden existir entre varias variables estadísticas. Para ello, es usual suponer que una de esas variables (que notaremos por y) depende, en algún sentido, de las demás variables (que notaremos por x_1, x_2, \dots, x_N). Se dice, entonces, que y es la variable *dependiente* (también llamada *de salida*, *endógena* o *respuesta*) y que $X = (x_1, \dots, x_N)$ representa las *variables independientes* (también llamadas *de entrada*, *exógenas* o *explicativas*). De esta forma, se trata de buscar una función que permita explicar los valores de y dependiendo de los valores de las variables x_1, x_2, \dots, x_N . Si sólo hay una variable explicativa (es decir, si $N = 1$), el problema de regresión se denomina *simple*, mientras que si hay más de una variable explicativa, hablamos de regresión *múltiple*.

De esta forma, cuando el conocimiento de una o varias variables (x_1, x_2, \dots, x_N) nos permite inferir, en mayor o menor grado, el valor de otra variable (y), se

dice que existe una relación de *dependencia estadística* (o *estocástica*) entre ellas. Los métodos de regresión estudian la construcción de un modelo estadístico (a través de una función) que explique esta relación de dependencia de la variable y respecto a las variables explicativas x_1, x_2, \dots, x_N . El objetivo esencial del análisis de regresión es el de predecir el comportamiento de la variable respuesta conocido el comportamiento de las variables explicativas que influyen sobre ella.

Formalmente, dado un conjunto de n datos $\{X_i, y_i\}_{i=1}^n$ obtenidos a partir de la variable aleatoria conjunta (X, y) (donde $X_i \in \mathbb{R}^N$ y suponemos que y_i es el valor de la variable respuesta y correspondiente a $X_i \in \mathbb{R}^N$) y una función $f(x, \underline{a})$, el objetivo es encontrar el vector de parámetros $\underline{a} = (a_1, a_2, \dots, a_k) \in \mathbb{R}^k$ tal que

$$y_i = f(X_i, \underline{a}) \quad \text{para cada } i \in \{1, 2, \dots, n\}.$$

2.1.2. Las funciones de pérdida

Para afrontar el problema de Regresión Clásico, se considera una función que mida la distancia entre el valor real y_i obtenido en el experimento y el valor $f(X_i, \underline{a})$ predicho por el modelo. Dicha forma de medir se denomina *función de pérdida*. La elección más habitual para la función de pérdida es la norma L_p , que se basa en las funciones

$$d_p(x, y) = |x - y|^p \quad \text{para cada } x, y \in \mathbb{R},$$

donde p es un número real positivo. En sentido estricto, veremos que d_p es una auténtica métrica únicamente si $0 < p \leq 1$ (si $p > 1$, entonces d_p es una semimétrica). De esta forma, el error cometido por el modelo puede expresarse como

$$L_p(y - f(X, \underline{a}), X) = \|y - f(X, \underline{a})\|^p \quad (2.1)$$

Los errores concretos de predicción que se obtienen en la práctica, es decir, los

valores $\|y_i - f(X_i, \underline{a})\|^p$, se denominan *residuos*. Si consideramos una muestra aleatoria de datos $\{X_i, y_i\}_{i=1}^n$ obtenida de (X, y) , y tomamos $p = 2$, la siguiente suma

$$SRC = \sum_{i=1}^n \|y_i - f(X_i, \underline{a})\|^2,$$

que nos aporta un valor global del error total cometido, se denomina *suma de residuos al cuadrado*. El objetivo del proceso de regresión es el de encontrar los valores óptimos para el vector de parámetros $\underline{a} = (a_1, a_2, \dots, a_k)$ de manera que la suma de residuos asociados a la muestra utilizando el criterio L_p sea mínima.

Cuando $p = 1$, el criterio L_1 de minimización de residuos se denomina *método de desviaciones absolutas*, mientras que si $p = 2$, el criterio L_2 de minimización de residuos cuadráticos es conocido como *método de mínimos cuadrados*. Cada uno de ellos tiene sus ventajas y sus inconvenientes. Desde luego, el criterio de mínimos cuadrados es el más utilizado en el ámbito de la regresión por sus magníficas propiedades y sus sencillez de uso. No obstante, el criterio L_1 es más robusto en el sentido de que es menos sensible a la existencia de valores *atípicos* (es decir, valores excepcionales de la distribución que son más grandes o más pequeños que la mayor parte de los valores de la muestra). De esta forma, el criterio L_1 puede ser útil en estudios donde la presencia de los valores extremos deba ser ignorada de manera segura y efectiva. Si fuese importante prestar atención a los valores atípicos, entonces el método de mínimos cuadrados sería una opción mejor. Las soluciones encontradas utilizando el método de mínimos cuadrados son *estables* en el sentido de que, para cualquier pequeño ajuste de los datos, el ajuste de regresión se moverá sólo ligeramente; es decir, los parámetros de regresión son funciones continuas dependientes de los datos y, además, este procedimiento siempre produce una única solución (en contraste con el método de desviaciones absolutas, el cual puede arrojar soluciones múltiples e inestables).

2.1.3. El método de mínimos cuadrados para la regresión lineal

Teniendo en cuenta las propiedades anteriores, la función de pérdida más utilizada es aquella que utiliza el criterio L_2 . En este caso, considerando una función de pérdida cuadrática, diremos que se realiza un *ajuste por mínimos cuadrados*. Otra de las suposiciones más frecuentes (que lleva a un modelo relativamente simple) consiste en asumir que la función f que se emplea para desarrollar el proceso de regresión es lineal (en caso de no serlo, en algunas ocasiones, el modelo se puede *linealizar* mediante transformaciones de las variables). La Regresión Lineal Clásica, para cada observación, asume el modelo lineal siguiente:

$$y_i = \beta_0 + \beta_1 x_{1i} + \dots + \beta_N x_{Ni} + \varepsilon_i \quad \text{para cada } i \in \{1, 2, \dots, n\},$$

donde los parámetros β_i son desconocidos y los residuos aleatorios ε_i han de cumplir las siguientes condiciones.

1. El valor esperado de los residuos es cero, es decir, $E[\varepsilon_i] = 0$, para cada $i \in \{1, 2, \dots, n\}$.
2. La varianza de los residuos es un valor constante, o sea, $Var(\varepsilon_i) = \sigma^2$ para cada $i \in \{1, 2, \dots, n\}$ (esta propiedad se conoce como *homocedasticidad*).
3. Hay una ausencia de autocorrelación entre los términos aleatorios de los distintos elementos de la muestra, lo que se traduce en que las covarianzas $Cov(\varepsilon_i, \varepsilon_j)$ toman el valor cero, para cualesquiera valores $i, j \in \{1, 2, \dots, n\}$ con $i \neq j$.

El modelo lineal general se puede expresar en notación matricial como sigue:

$$y = X\beta + \varepsilon,$$

donde y es el vector de dimensión n que contiene las observaciones de la variable respuesta, X es la matriz de dimensión $n \times (N+1)$ que contiene las observaciones de la variable explicativa y una columna de unos, y β es el vector $(N+1) \times 1$ que contiene los parámetros desconocidos que queremos estimar, es decir:

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1N} \\ 1 & x_{21} & \cdots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nN} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_N \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

Para encontrar una solución al problema podemos considerar una función pérdida como la función dada en (2.1). Sin embargo, tal y como hemos comentado al principio, el método de estimación por mínimos cuadrados es el más utilizado y consiste en minimizar la suma de los residuos al cuadrado, es decir, minimizar la expresión

$$SRC = \|\varepsilon\|^2 = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon^t \varepsilon = (y - X\beta)^t (y - X\beta) = y^t y - 2y^t X\beta + \beta^t (X^t X) \beta.$$

Derivando con respecto a β e igualando a cero, obtenemos que el valor óptimo para el parámetro $(N+1)$ -dimensional β es:

$$\hat{\beta} = (X^t X)^{-1} X^t y. \quad (2.2)$$

Este valor se conoce como *estimador de mínimos cuadrados*. El teorema de Gauss-Markov asegura que éste es el estimador insesgado de mínima varianza (lo que se conoce como *eficiente*). Esta propiedad es una de las razones por las que este método es ampliamente utilizado en diferentes contextos. Sin embargo, este resultado no garantiza que la varianza de los estimadores sea necesariamente pequeña. En concreto, el teorema de Gauss-Markov asegura que los estimadores mínimo-cuadráticos son los mejores dentro de la clase de estimadores que son insesgados y que se obtienen como funciones lineales de las observaciones, pero no asegura que no puedan existir otros estimadores (por ejemplo, no lineales)

que sean mejores. Por consiguiente, podría existir un estimador sesgado que tuviera menor varianza que el calculado por mínimos cuadrados.

Cuando la varianza σ^2 es desconocida, el estimador que se emplea para aproximarla es

$$\hat{\sigma}^2 = \frac{1}{n + N - 1} \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

donde los valores \hat{y}_i se obtienen utilizando la solución dada en (2.2). A la raíz cuadrada de $\hat{\sigma}^2$ se le denomina, a veces, el *error estándar de regresión* y tiene las mismas unidades que la variable respuesta. Dicha medida representa la desviación estándar de los valores de y sobre la recta de regresión estimada y puede utilizarse como indicador del grado de ajuste del modelo de regresión, aunque al tener las mismas unidades de medida que la variable endógena no nos permite comparar la bondad de ajuste de dos modelos con variable endógena diferente. Además, debido a que $\hat{\sigma}^2$ depende de los residuos del modelo, cualquier violación de las hipótesis que se asumen sobre el modelo o cualquier especificación errónea del mismo pueden influir en $\hat{\sigma}^2$ como estimador de σ^2 . Por esta razón, decimos que $\hat{\sigma}^2$ es un estimador de σ^2 dependiente del modelo.

Utilizando la ecuación lineal del modelo y la estimación calculada para el vector de parámetros desconocidos, podemos hacer estimaciones puntuales (o predicciones) para la variable y , que denotamos por \hat{y}_i , para un valor determinado X_i de X , mediante la expresión:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \dots + \hat{\beta}_N x_{Ni} = X \hat{\beta}.$$

Dado que cada estimación está sujeta a un margen de error, puede ser conveniente determinar un intervalo de confianza en el que suponemos que está contenido el verdadero valor de la variable estimada. Para obtener estos intervalos, para resolver contrastes de hipótesis y para estudiar el comportamiento de los parámetros que intervienen en el modelo, se supone que los residuos siguen una determinada distribución. Por lo general, se supone que dichos errores siguen

una distribución Normal estándar:

$$\varepsilon_i \rightsquigarrow N(0, 1).$$

En estas condiciones, el estimador para el vector de coeficientes del modelo sigue otra distribución Normal, cuyos parámetros son:

$$\hat{\beta} \rightsquigarrow N\left(\beta, \sigma^2 (X^t X)^{-1}\right).$$

A partir de esta distribución, es posible obtener un intervalo de confianza al $(1 - \alpha) 100$ por ciento para los valores de β y resolver contrastes de hipótesis sobre este parámetro concreto. También hemos de destacar el hecho de que, para resolver un problema de regresión, es necesario un número mínimo de observaciones de tal forma que, a medida que este número aumenta, también aumentan los grados de libertad y, así, los intervalos de confianza se hacen cada vez más precisos.

2.1.4. Medidas de bondad del ajuste

Una primera evaluación del ajuste de una recta de regresión puede hacerse mediante la varianza residual, considerando que el modelo ajustado será poco representativo cuando la varianza residual es grande. Sin embargo, esta medida no es útil para comparar rectas de regresión de variables distintas, porque depende de las unidades de medida de la variable dependiente.

Como medida más adecuada del ajuste, existe un indicador de la dispersión relativa que expresa la proporción o el grado de variabilidad de la variable dependiente explicada por el modelo lineal ajustado. A este indicador se le denomina *coeficiente de determinación* y resuelve el problema de las unidades de medida. Dicho coeficiente se representa por R^2 o r^2 . Para su cálculo, tenemos en cuenta que la variabilidad del conjunto de datos se mide a través de diferentes sumas de cuadrados: la *suma total de cuadrados*, STC ,

$$STC = \sum_{i=1}^n \|y_i - \bar{y}\|^2,$$

la suma de cuadrados de regresión (también llamada *suma de cuadrados explicada*), SEC ,

$$SEC = \sum_{i=1}^n \| f(X_i, \underline{a}) - \bar{y} \|^2,$$

y la suma de cuadrados de los residuos (también llamada *suma residual*), SRC ,

$$SRC = \sum_{i=1}^n \| y_i - f(X_i, \underline{a}) \|^2.$$

La definición más general que se puede dar del coeficiente de determinación es la siguiente:

$$R^2 = 1 - \frac{SRC}{STC}.$$

En algunos casos (por ejemplo, para los modelos lineales), la suma total de cuadrados es igual a la suma de las otras dos sumas de cuadrados y la definición anterior de R^2 es equivalente a

$$R^2 = \frac{SEC}{STC} = \frac{SEC/n}{STC/n}.$$

En esta última expresión, este coeficiente se obtiene como el cociente entre la varianza explicada por la regresión y la varianza total y, por ello, a menudo se interpreta como la proporción de variabilidad de la respuesta explicada por la regresión. En este caso, el coeficiente de determinación tiene la propiedad de que sólo toma valores comprendidos entre 0 y 1 ($0 \leq R^2 \leq 1$), y es usual expresarlo en porcentaje. Puede comprobarse que si la recta ajustada pasa por todos los puntos observados, entonces la varianza residual es nula y $R^2 = 1$, en cuyo caso el modelo de regresión explica el 100% de la variabilidad de la variable y . Así, valores de R^2 cercanos a 1 se interpretan como que la mayor parte de la variabilidad de la respuesta es explicada por el modelo de regresión. Por otro lado, si la varianza explicada por la regresión es nula, entonces $R^2 = 0$. Esto sucede cuando la variable independiente (X) no es capaz de explicar ninguna variación de la variable dependiente (y).

El coeficiente de correlación lineal de Pearson (r_{xy}) tiene por objeto medir el grado de asociación lineal entre dos variables. Este coeficiente verifica que

$r_{xy}^2 = R^2$, por lo que toma valores entre -1 y 1 . Sin embargo, posee la ventaja de que su signo aporta nueva información: si $r = -1$, hay una asociación lineal negativa perfecta entre las variables de forma que cuando una variable crece, la otra decrece; por contra, si $r = 1$, hay una asociación lineal positiva perfecta entre las variables, y el crecimiento de una variable conlleva también aumento en la segunda; finalmente, si $r = 0$, no hay asociación lineal entre las variables. Es importante tener en cuenta que el coeficiente de correlación lineal mide el grado de asociación lineal entre las variables, por lo que es un estadístico inapropiado si la relación no es lineal.

Si los parámetros se estiman mediante el método de mínimos cuadrados, el coeficiente R^2 aumentará a medida que crece el número de variables involucradas en el modelo. Sin embargo, disponer de más variables explicativas no siempre significa aumentar la precisión con la que podemos determinar la variable respuesta. Este hecho nos lleva a introducir un coeficiente alternativo, denominado *R^2 -ajustado* y denotado por \bar{R}^2 , que trata de eliminar la influencia de variables superfluas en la determinación de la bondad del ajuste, que se define como:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{n - 1}{n - N - 1}.$$

2.1.5. Diagnósis en el modelo de regresión lineal

Un aspecto que se olvida frecuentemente es que los modelos de regresión se basan en hacer unas determinadas suposiciones sobre los datos y que éstas no siempre se cumplen, por lo que es preciso comprobar si las hipótesis básicas del modelo se dan en nuestros datos. Es lo que se conoce como **diagnósis del modelo**. Tendremos que contrastar entre otras las hipótesis de linealidad, normalidad, homocedasticidad, no autocorrelación e independencia.

Hipótesis de normalidad: La hipótesis de normalidad de los errores aleatorios es necesaria para realizar inferencias respecto a los parámetros y

para la construcción de los intervalos de predicción. El supuesto de normalidad no es absolutamente esencial si el objetivo planteado consistiese únicamente en la estimación puntual de los parámetros del modelo. Como hemos visto, los estimadores mínimo cuadráticos son estimadores óptimos, con independencia de la distribución de probabilidad que siga el término de perturbación aleatoria. Si además la variable aleatoria se distribuye normalmente, puede demostrarse que los estimadores mínimo cuadráticos tenderán a serlo también, a medida que aumente indefinidamente el tamaño de la muestra.

Hipótesis de no multicolinealidad: Esta hipótesis implica que ninguna de las variables explicativas puede obtenerse como combinación lineal de las demás. Si una combinación lineal de las columnas de X puede ser cercana a 0, el cálculo de $(X^t X)^{-1}$ se vuelve inestable y esto produce un aumento considerable del valor absoluto de los coeficientes.

Hipótesis de homocedasticidad: El término σ^2 puede no ser constante, y variar entre las diversas observaciones, lo que se denomina problema de *heterocedasticidad*. En presencia de heterocedasticidad los estimadores mínimos cuadráticos de los coeficientes de regresión siguen siendo insesgados aunque su varianza ya no será la mínima posible. Esto implica que los coeficientes de regresión tendrán errores estándares más grandes y los contrastes se verán gravemente afectados debido a una pérdida de sensibilidad.

Hipótesis de no autocorrelación e independencia: La acepción más frecuente del término autocorrelación hace referencia a la correlación existente entre los elementos de una serie de observaciones ordenadas en el tiempo o en el espacio. En el contexto de la regresión se supone que la autocorrelación no existe en los errores aleatorios. Si esto no fuera así, estaríamos ante el problema de existencia de autocorrelación. Los efectos de

la autocorrelación sobre los estimadores pueden ser muy graves: los intervalos de confianza y contrastes de hipótesis basados en las distribuciones t de student y F de Snedecor ya no serían apropiados. Tampoco lo serían los intervalos para las predicciones. Generalmente, la subestimación de la varianza proporciona al analista una falsa impresión de precisión.

Hipótesis de linealidad: La hipótesis de relación lineal entre la variable respuesta y las variables explicativas es una asunción básica en el análisis de regresión. En ocasiones encontramos que esta suposición es inapropiada.

2.1.6. Selección de variables explicativas

En muchas situaciones se dispone de un conjunto grande de posibles variables explicativas, una primera pregunta es saber si todas las variables deben de entrar en el modelo de regresión y, en caso negativo, se quiere saber qué variables deben entrar y qué variables no deben entrar en el modelo de regresión. Intuitivamente parece bueno introducir en el modelo todas las variables explicativas significativas (según el contraste individual de la t) al ajustar el modelo con todas las variables posibles. Pero este procedimiento no es adecuado porque en la varianza del modelo influye el número de variables del modelo y ésta crece al aumentar el número de variables explicativas. Además puede haber problemas de multicolinealidad cuando hay muchas variables explicativas.

Para responder a este problema se dispone de diferentes procedimientos estadísticos. Bajo la hipótesis de que la relación entre las variables regresoras y la variable respuesta es lineal existen procedimientos *paso a paso* (o *stepwise*) que permiten elegir el subconjunto de variables regresoras que deben estar en el modelo. El algoritmo *paso a paso* tiene las ventajas del algoritmo de introducción progresiva pero lo mejora al no mantener fijas en el modelo las variables que ya entraron en una etapa, evitando de esta forma problemas de multicolinealidad. En la práctica, es un algoritmo muy utilizado ya que proporciona resultados

razonables cuando se tiene un número grande de variables explicativas.

Finalmente comentar que cuando tenemos una variable dependiente dicotómica que deseamos predecir, o para la que queremos evaluar la asociación o relación con otras (más de una) variables independientes y de control, el procedimiento a realizar es una *Regresión Logística (binaria) Multivariante*. La Regresión Logística es probablemente el tipo de análisis multivariante más empleado en Ciencias de la Vida. No cabe ninguna duda que este tipo de regresión es una de las herramientas estadísticas con mejor capacidad para el análisis de datos en investigación clínica y epidemiología, de ahí su amplia utilización. En este caso, la variable no sigue una distribución normal, sino binomial. Para dar solución a este problema se hace una transformación del modelo estándar, resultando el modelo logístico

$$\pi(x) = \frac{e^{X\beta}}{1 + e^{X\beta}}$$

El método de estimación que se utiliza en este caso es el de *máxima verosimilitud* que consiste en elegir los parámetros de un modelo como aquellos que hagan máxima la probabilidad de obtener la muestra observada, de acuerdo con los datos disponibles.

2.2. Teoría de Conjuntos Difusos

Desde la aparición de la teoría de conjuntos difusos, son incontables las aplicaciones que se han hecho de ella en el mundo de la investigación y de la ingeniería. A continuación, presentamos algunos conceptos fundamentales de esta teoría.

2.2.1. Valores Difusos

En esta sección vamos a introducir las nociones que utilizaremos de conjunto difuso y de número difuso. Para ello, recordamos brevemente la motivación inicial que llevó a Zadeh, en 1969, a definir el concepto de conjunto difuso.

Sea X un conjunto no vacío que sirva a modo de universo. Un subconjunto A de X puede verse como una colección concreta de elementos de X . Dado un elemento x de X , escribimos $x \in A$ si x está incluido en A y escribimos $x \notin A$ si el elemento x no pertenece a A . De esta forma, a cada subconjunto A de X podemos asociarle la siguiente función:

$$\mu_A : X \rightarrow \{0, 1\}, \quad \mu_A(x) = \begin{cases} 1, & \text{si } x \in A, \\ 0, & \text{si } x \notin A. \end{cases}$$

La función μ_A se denomina *función característica del subconjunto A* y, dentro de X , lo determina completamente puesto que un elemento de X pertenece a A si μ_A toma el valor 1 sobre él, y no pertenece a A si μ_A toma el valor cero. Si reemplazamos el conjunto discreto $\{0, 1\}$ por el intervalo real $[0, 1]$, una aplicación $F : X \rightarrow [0, 1]$ puede expresar, para cada punto x del universo X , la probabilidad (o certidumbre) que tenemos de que el elemento x pertenezca al conjunto difuso. De esta forma, un conjunto difuso no es más que una colección de elementos en un universo provistos de un cierto grado (expresado entre 0 y 1) de pertenencia al conjunto. El grado máximo de pertenencia es 1 cuando

estamos seguros de que el elemento pertenece al conjunto, y el grado mínimo es 0 cuando tenemos la seguridad de que el elemento no pertenece al conjunto. No obstante, existe una amplia gama de valores intermedios que expresan nuestro grado de certeza sobre la pertenencia o no del elemento al conjunto.

Existen muchas nociones diferentes de número difuso, en el que el máximo de la función de pertenencia puede ser menor que 1 o en la que sus conjuntos de nivel no son necesariamente intervalos (véase [65, 66, 49, 50]). No obstante, para nuestros propósitos, seguiremos la siguiente definición. En lo que sigue, \mathbb{R} denotará el conjunto de los números reales, $\mathbb{R}_0^+ = [0, \infty)$ será el intervalo formado por todos los números reales no negativos e \mathbb{I} será el intervalo cerrado $[0, 1]$.

Definición 2.2.1 (Conjunto difuso) *Un conjunto difuso sobre \mathbb{R} es una función $\mathcal{A} : \mathbb{R} \rightarrow \mathbb{I}$.*

Como hemos indicado anteriormente, el valor $\mathcal{A}(x)$ denota el grado de pertenencia del número real x al conjunto difuso \mathcal{A} .

Definición 2.2.2 (α -corte) *Dado un conjunto difuso \mathcal{A} y un número $\alpha \in (0, 1]$, el α -corte (o el conjunto de nivel α) de \mathcal{A} , que denotamos por \mathcal{A}_α , es el conjunto*

$$\mathcal{A}_\alpha = \{x \in \mathbb{R} : \mathcal{A}(x) \geq \alpha\}.$$

Los números difusos que utilizaremos son conjuntos difusos que cumplen ciertas condiciones (para más detalles acerca de la noción de número difuso, remitimos al lector a [17, 18, 19, 40, 68]).

Definición 2.2.3 (Número difuso) *Un número difuso sobre \mathbb{R} es un conjunto difuso \mathcal{A} sobre \mathbb{R} tal que, para todo $\alpha \in (0, 1]$, el conjunto de nivel $\mathcal{A}_\alpha = \{x \in \mathbb{R} : \mathcal{A}(x) \geq \alpha\}$ es un intervalo cerrado y no vacío de \mathbb{R} .*

Es claro que si $0 < \alpha \leq \beta \leq 1$, entonces $\mathcal{A}_\beta \subseteq \mathcal{A}_\alpha$. De esta forma, \mathcal{A}_1 es siempre el menor conjunto de nivel del número difuso \mathcal{A} .

Definición 2.2.4 (Núcleo) *El núcleo de un número difuso \mathcal{A} es su 1-corte, es decir,*

$$\ker \mathcal{A} = \mathcal{A}_1 = \{x \in \mathbb{R} : \mathcal{A}(x) = 1\}.$$

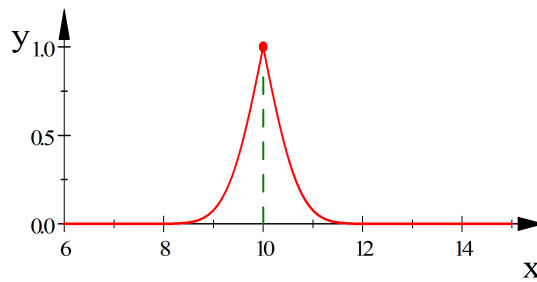


Figura 2.1: Un ejemplo de número difuso cuyo núcleo es $\{10\}$

Dado que sólo consideramos números difusos cuyo núcleo sea no vacío, éstos deben cumplir la hipótesis de *normalidad*, es decir, para cada número difuso \mathcal{A} , existe $x_0 \in X$ tal que $\mathcal{A}(x_0) = 1$. Sin embargo, con la noción anterior, un número difuso \mathcal{A} podría no tomar el valor cero (véase la Figura 2.2).

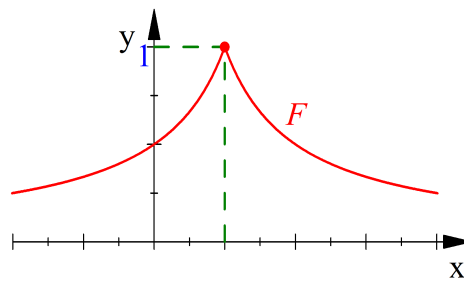


Figura 2.2: Un número difuso que no se anula en \mathbb{R} .

Tiene entonces sentido realizar la siguiente definición.

Definición 2.2.5 (Soporte) *El soporte de un número difuso \mathcal{A} es la clausura en \mathbb{R} del conjunto de valores sobre los que \mathcal{A} toma valores estrictamente positivos,*

es decir,

$$\text{supp } \mathcal{A} = \overline{\{x \in \mathbb{R} : \mathcal{A}(x) > 0\}}.$$

En lo que sigue, sólo consideraremos números difusos con soporte compacto y denotaremos por \mathcal{F} a la familia de todos los números difusos (de soporte compacto). De esta forma, los números difusos sobre \mathbb{R} que utilizaremos son aquéllos cuyos α -cortes son intervalos cerrados y cuyo soporte es acotado. En particular, para cada $\alpha \in \mathbb{I}$, el α -corte \mathcal{A}_α de un número difuso \mathcal{A} es un subintervalo compacto (y no vacío) de \mathbb{R} . Consecuentemente, puede expresarse como $\mathcal{A}_\alpha = [\underline{a}_\alpha, \bar{a}_\alpha]$, donde \underline{a}_α es el extremo inferior y \bar{a}_α es el extremo superior del intervalo \mathcal{A}_α . Siguiendo esta notación, también denotamos el soporte de \mathcal{A} por $\text{supp } \mathcal{A} = \mathcal{A}_0 = [\underline{a}_0, \bar{a}_0]$.

Dado un número difuso \mathcal{A} y variando α on \mathbb{I} , podemos considerar dos funciones $\underline{a}, \bar{a} : \mathbb{I} \rightarrow \mathbb{R}$ donde $\underline{a}(\alpha) = \underline{a}_\alpha$ y $\bar{a}(\alpha) = \bar{a}_\alpha$ para todo $\alpha \in \mathbb{I}$. Estas funciones determinan completamente el número difuso \mathcal{A} , es decir, es equivalente conocer un número difuso $\mathcal{A} : \mathbb{R} \rightarrow \mathbb{I}$ que sus funciones asociadas $\underline{a}, \bar{a} : \mathbb{I} \rightarrow \mathbb{R}$.

Lema 2.2.6 (Véase [24]) Un conjunto difuso \mathcal{A} es un número difuso si, y solo si, existen dos funciones continuas a la izquierda $\underline{a}, \bar{a} : \mathbb{I} \rightarrow \mathbb{R}$ tales que \underline{a} es no decreciente, \bar{a} es no creciente y $\mathcal{A}_\alpha = [\underline{a}_\alpha, \bar{a}_\alpha]$ para todo $\alpha \in \mathbb{I}$.

Un intervalo compacto real queda determinado por su *centro* (su punto medio) y su *amplitud* (la distancia entre sus extremos). Utilizaremos el término *radio* de un intervalo compacto para denotar a la mitad de su amplitud, es decir, a la distancia entre su centro y uno de sus extremos. De esta forma, si $c \in \mathbb{R}$ es el centro y $r \geq 0$ es el radio, entonces el intervalo es $[c - r, c + r]$. Aplicadas estas nociones al núcleo $\mathcal{A}_1 = [\underline{a}_1, \bar{a}_1]$ de un número difuso \mathcal{A} , introducimos los siguientes conceptos.

Definición 2.2.7 (Centro y amplitudes) *El número real*

$$\mathcal{D}_c \mathcal{A} = \frac{\underline{a}_1 + \bar{a}_1}{2} \tag{2.3}$$

es el centro de un número difuso \mathcal{A} , y su amplitud (en inglés, spread) es

$$\text{spr } \mathcal{A} = \frac{\bar{a}_1 - \underline{a}_1}{2}.$$

Nótese que podemos considerar dos funciones $\mathcal{D}_c : \mathcal{F} \rightarrow \mathbb{R}$ y $\text{spr} : \mathcal{F} \rightarrow \mathbb{R}_0^+$ que asignan, a cada número difuso \mathcal{A} , su centro y su radio, respectivamente. Como toman valores reales, ambas funciones pierden información del número difuso. En este sentido, se las denomina *defusificaciones*.

En lo que sigue, una *defusificación*¹ será un proceso en el que las cantidades difusas se aproximan por números reales (vistos como números difusos escalares o *crisp*), es decir, una función $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$. Existen muchos *métodos de valoración* diferentes con una amplia variedad de nombres, algunos muy conocidos: *valor medio*, *ambigüedad*, *valor esperado*, *centro de gravedad*, *área bajo el número difuso*, etc. (véase, por ejemplo, [5, 66]). Nosotros no impondremos condiciones adicionales a las defusificaciones que consideraremos en esta memoria. No obstante, acorde a (2.3), una de las defusificaciones más importantes (denotada por \mathcal{D}_c) es la que a cada número difuso \mathcal{A} le asocia su centro $\mathcal{D}_c \mathcal{A}$.

2.2.2. Algunas clases de números difusos

A continuación, presentamos una familia de números difusos que poseen formas geométricas muy simples pero que, a la vez, son ampliamente utilizados en la investigación actual ya que satisfacen las necesidades de modelización que se plantean en los problemas difusos y, en general, son más fáciles de fijar y de manejar.

Definición 2.2.8 (Número difuso generalizado de tipo LR generalizado) *Un número difuso (generalizado) de tipo LR (left right, de Dubois y Prade) [17,*

¹Del inglés *defuzzification*. A lo largo de esta Memoria, preferimos el término *defusificación* en lugar de *defuzzificación* porque nos parece más natural.

18, 19] es un número difuso $\mathcal{A} = (a_1/a_2/a_3/a_4)_{LR}$, donde $a_1, a_2, a_3, a_4 \in \mathbb{R}$, $a_1 \leq a_2 \leq a_3 \leq a_4$, dado por la siguiente función de pertenencia:

$$\mathcal{A}(x) = \begin{cases} L\left(\frac{x - a_1}{a_2 - a_1}\right), & \text{si } a_1 < x < a_2, \\ 1, & \text{si } a_2 \leq x \leq a_3, \\ R\left(\frac{a_4 - x}{a_4 - a_3}\right), & \text{si } a_3 < x < a_4, \\ 0, & \text{en cualquier otro caso,} \end{cases}$$

donde $L, R : \mathbb{I} \rightarrow \mathbb{I}$ son funciones continuas y estrictamente crecientes tales que $L(0) = R(0) = 0$ y $L(1) = R(1) = 1$.

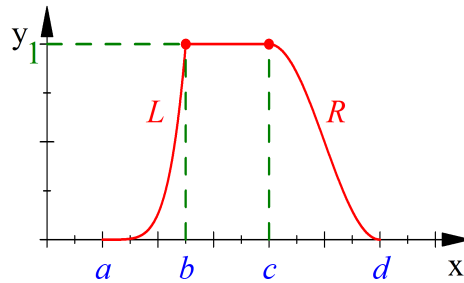


Figura 2.3: Representación gráfica del número difuso LR $(a/b/c/d)_{LR}$.

Claramente, el núcleo de \mathcal{A} es $[a_2, a_3]$ y su soporte es $[a_1, a_4]$. Una clase de números difusos (generalizados) de tipo LR que están determinados de manera única por sus núcleos y sus soportes son los números difusos trapezoidales.

Definición 2.2.9 (Número difuso trapezoidal (generalizado)) *Un número difuso (generalizado) trapezoidal es un número difuso $\mathcal{A} = (A^I/A^J/A^R/A^S)$, donde $A^I, A^J, A^R, A^S \in \mathbb{R}$ son números reales que verifican $A^I \leq A^J \leq A^R \leq A^S$, y*

cuya función de pertenencia viene dada por:

$$\mathcal{A}(x) = \begin{cases} \frac{x - A^I}{A^J - A^I}, & \text{si } A^I < x < A^J, \\ 1, & \text{si } A^J \leq x \leq A^R, \\ \frac{A^S - x}{A^S - A^R}, & \text{si } A^R < x < A^S, \\ 0, & \text{en cualquier otro caso.} \end{cases}$$

Denotaremos por \mathcal{T} a la familia formada por todos los números difusos trapezoidales (generalizados) sobre \mathbb{R} .

La Figura 2.4 muestra la representación gráfica del número difuso trapezoidal $(a/b/c/d)$, donde a , b , c y d son números reales tales que $a < b < c < d$.

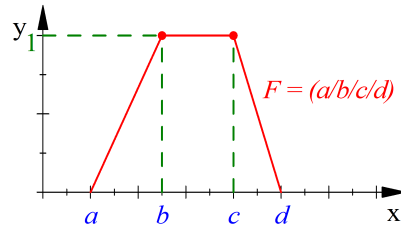


Figura 2.4: Representación gráfica del número difuso trapezoidal $(a/b/c/d)$.

Los números difusos trapezoidales son casos particulares de los números difusos LR, donde $L(x) = R(x) = x$ para todo $x \in \mathbb{I}$. Claramente, el núcleo de un número trapezoidal \mathcal{A} es $[A^J, A^R]$, y su soporte es $[A^I, A^S]$. De esta forma, comprobamos que estos números están caracterizados por sus núcleos y sus soportes, es decir, si \mathcal{A} y \mathcal{B} son dos números difusos trapezoidales, entonces $\mathcal{A} = \mathcal{B}$ si, y sólo si, $\ker \mathcal{A} = \ker \mathcal{B}$ y $\text{supp } \mathcal{A} = \text{supp } \mathcal{B}$.

En realidad, el concepto tradicional de número difuso trapezoidal ocurre cuando $A^I < A^J < A^R < A^S$, pues su representación gráfica es la de un trapecio de altura 1, cuya base menor es su núcleo $[A^J, A^R]$ y cuya base mayor es su soporte $[A^I, A^S]$. En este caso, los números A^I , A^J , A^R y A^S se denominan

esquinas (en inglés, *corners*) del número difuso. No obstante, permitiendo la igualdad entre algunos de estos valores, incluimos algunas clases de números difusos que no son continuos como aplicaciones.

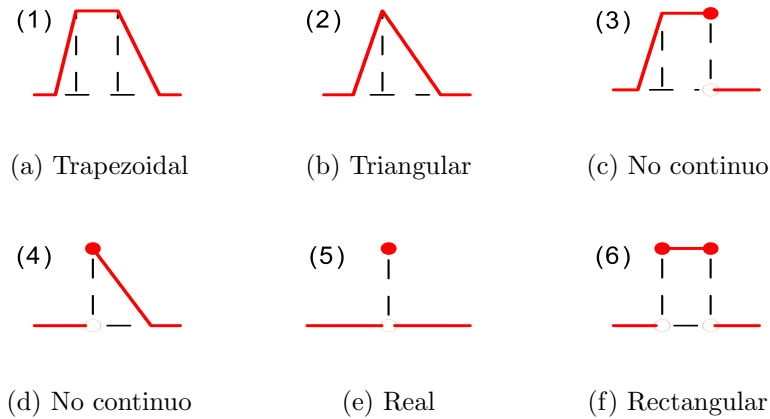


Figura 2.5: Distintas clases de números difusos trapezoidales (generalizados).

Basándonos en la forma de su representación gráfica, podemos distinguir entre los siguientes tipos de números trapezoidales.

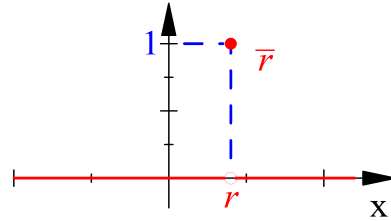
Definición 2.2.10 *Un número difuso trapezoidal $\mathcal{A} = (A^I/A^J/A^R/A^S)$ es:*

- real (o crisp) si $A^I = A^J = A^R = A^S$;
- rectangular si $A^I = A^J \leq A^R = A^S$;
- triangular si $A^I \leq A^J = A^R \leq A^S$.

Claramente, \mathbb{R} puede embeberse en \mathcal{T} si interpretamos cada número real como su correspondiente valor difuso *crisp*: si $r \in \mathbb{R}$, entonces $\tilde{r} \in \mathcal{T}$, donde

$\tilde{r} : \mathbb{R} \rightarrow \mathbb{I}$ está dado, para todo $x \in \mathbb{R}$, por

$$\tilde{r}(x) = \begin{cases} 1, & \text{si } x = r, \\ 0, & \text{si } x \neq r. \end{cases}$$



El *centro* del número trapezoidal $\mathcal{A} = (A^I/A^J/A^R/A^S)$ es

$$A^c = \frac{A^J + A^R}{2}.$$

No obstante, en este caso, cabe hablar de sus tres *amplitudes* (o *spreads*), que vienen dados por los siguientes valores no negativos:

$$A^m = \frac{A^R - A^J}{2} \geq 0, \quad A^\ell = A^J - A^I \geq 0, \quad A^s = A^S - A^R \geq 0.$$

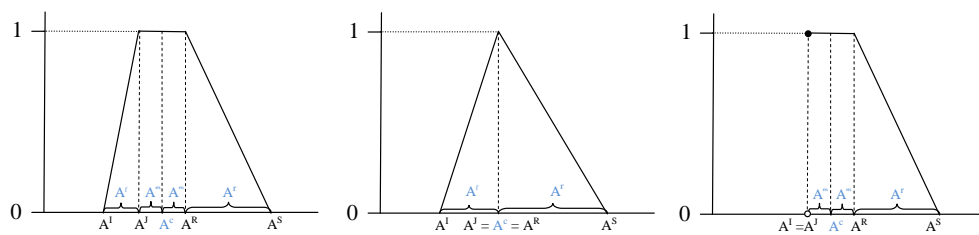


Figura 2.6: Amplitudes de números trapezoidales generalizados.

El centro y las tres amplitudes también determinan un número difuso trapezoidal ya que

$$\begin{aligned} A^I &= A^c - A^m - A^\ell, \\ A^J &= A^c - A^m, \\ A^R &= A^c + A^m, \\ A^S &= A^c + A^m + A^s. \end{aligned}$$

En este sentido, es usual utilizar la siguiente notación:

$$\mathcal{A} = \text{Tra}(A^c, A^m, A^\ell, A^s), \quad (2.4)$$

donde el primer argumento A^c denota su centro y sus tres últimos argumentos son las amplitudes del número difuso trapezoidal \mathcal{A} (véase la Figura 2.6).

Obsérvese que un número difuso trapezoidal $\mathcal{A} = \text{Tra}(A^c, A^m, A^\ell, A^s)$ es triangular si, y sólo si, $A^m = 0$. De esta forma, denotaremos por

$$\mathcal{A} = \text{Tri}(A^c, A^\ell, A^s)$$

al único número difuso triangular cuyo centro es A^c y cuyas amplitudes (izquierda y derecha) son A^ℓ y A^s . Y en el caso de que las amplitudes izquierda y derecha sean iguales, denotaremos por

$$\mathcal{A} = \text{Tri}(A^c, A)$$

al único número difuso triangular simétrico cuyo centro es A^c y cuyas amplitudes (izquierda y derecha) son $A^\ell = A^s = A$.

2.2.3. Operaciones entre números difusos

Existe una aritmética entre los números difusos (véase [40]) que extiende la aritmética real. Sin embargo, la forma más habitual para operar con números difusos es utilizando la aritmética de intervalos con los α -cortes y las funciones

\underline{a} y \bar{a} . De esta forma, la suma y el producto por escalares positivos se define como sigue: para cada $\mathcal{A}, \mathcal{B} \in \mathcal{F}$ y cada $r > 0$,

$$\begin{aligned} \mathcal{C} = \mathcal{A} + \mathcal{B} \quad \text{siendo} \quad \underline{c}_\alpha &= \underline{a}_\alpha + \underline{b}_\alpha \text{ y } \bar{c}_\alpha = \bar{a}_\alpha + \bar{b}_\alpha, \\ \mathcal{E} = r\mathcal{A} \quad \text{siendo} \quad \underline{e}_\alpha &= r\underline{a}_\alpha \text{ y } \bar{e}_\alpha = r\bar{a}_\alpha. \end{aligned}$$

Cuando $r \in \mathbb{R}$ y $\mathcal{A} \in \mathcal{F}$, escribiremos $r + \mathcal{A} = \tilde{r} + \mathcal{A}$.

- Proposición 2.2.11**
1. Si $a, b \in \mathbb{R}$ y $\mathcal{A} \in \mathcal{F}$, entonces $a + (b + \mathcal{A}) = (a + b) + \mathcal{A}$.
 2. Si $r \in \mathbb{R}$, $\mathcal{A}, \mathcal{B} \in \mathcal{F}$ y $r + \mathcal{A} = r + \mathcal{B}$, entonces $\mathcal{A} = \mathcal{B}$.
 3. Si $r \in \mathbb{R}$, $\mathcal{A}, \mathcal{B} \in \mathcal{F}$ y $\mathcal{A} = r + \mathcal{B}$, entonces $\mathcal{B} = (-r) + \mathcal{A}$.

DEMOSTRACIÓN: **(1)** Dado que $[\tilde{r}]_\alpha = [r, r]$ para cada $r \in \mathbb{R}$, deducimos que, para cada $\alpha \in \mathbb{I}$,

$$[r + \mathcal{A}]_\alpha = [\underline{r + \mathcal{A}}_\alpha, \overline{r + \mathcal{A}}_\alpha] = [r + \underline{a}_\alpha, r + \bar{a}_\alpha].$$

De esta forma,

$$\begin{aligned} [a + (b + \mathcal{A})]_\alpha &= [a + (b + \underline{a}_\alpha), a + (b + \bar{a}_\alpha)] = [(a + b) + \underline{a}_\alpha, (a + b) + \bar{a}_\alpha] \\ &= [(a + b) + \mathcal{A}]_\alpha. \end{aligned}$$

Dado que los α -cortes determinan el número difuso, concluimos que $a + (b + \mathcal{A}) = (a + b) + \mathcal{A}$.

- (2)** Teniendo en cuenta que, para cada $\alpha \in \mathbb{I}$,

$$[r + \underline{a}_\alpha, r + \bar{a}_\alpha] = [r + \mathcal{A}]_\alpha = [r + \mathcal{B}]_\alpha = [r + \underline{b}_\alpha, r + \bar{b}_\alpha],$$

concluimos que $\underline{a}_\alpha = \underline{b}_\alpha$ y $\bar{a}_\alpha = \bar{b}_\alpha$, por lo que $\mathcal{A} = \mathcal{B}$.

- (3)** Si, para cada $\alpha \in \mathbb{I}$,

$$[\underline{a}_\alpha, \bar{a}_\alpha] = [\mathcal{A}]_\alpha = [r + \mathcal{B}]_\alpha = [r + \underline{b}_\alpha, r + \bar{b}_\alpha],$$

entonces $\underline{a}_\alpha = r + \underline{b}_\alpha$ y $\bar{a}_\alpha = r + \bar{b}_\alpha$. De esta forma, $\underline{b}_\alpha = (-r) + \underline{a}_\alpha$ y $\bar{b}_\alpha = (-r) + \bar{a}_\alpha$, lo que, razonando como antes, significa que $\mathcal{B} = (-r) + \mathcal{A}$. ■

Lema 2.2.12 Si \mathcal{D} es una defusificación y $\mathcal{A} \in \mathcal{T}$, entonces existe un único $\mathcal{D}^* \mathcal{A} \in \mathcal{T}$ tal que $\mathcal{A} = \mathcal{D}\mathcal{A} + \mathcal{D}^* \mathcal{A}$.

DEMOSTRACIÓN: Sólo hay que tomar $a = \mathcal{D}\mathcal{A} \in \mathbb{R}$ y $\mathcal{D}^* \mathcal{A} = (-a) + \mathcal{A}$. Entonces, utilizando el primer apartado de la Proposición 2.2.11, se tiene que

$$\mathcal{D}\mathcal{A} + \mathcal{D}^* \mathcal{A} = a + ((-a) + \mathcal{A}) = (a - a) + \mathcal{A} = \mathcal{A}.$$

Y su unicidad se deduce del tercer apartado de la misma proposición. ■

En el siguiente resultado, describimos algunas relaciones entre esta aritmética y una familia de defusificaciones que incluye, como caso particular, a la defusificación \mathcal{D}_c .

Lema 2.2.13 Para todo $i \in \{1, 2, \dots, n\}$, sean $\beta_i, \gamma_i \in \mathbb{R}$ y $\alpha_i \in \mathbb{I}$ constantes arbitrarias y sea $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$ la defusificación dada por:

$$\mathcal{D}\mathcal{A} = \sum_{i=1}^n (\beta_i \underline{a}_{\alpha_i} + \gamma_i \bar{a}_{\alpha_i}) \quad \text{para cada } \mathcal{A} \in \mathcal{F}.$$

Entonces, para cada $\mathcal{A}, \mathcal{B} \in \mathcal{F}$ y cada $r > 0$, se tiene que:

$$\mathcal{D}(\mathcal{A} + \mathcal{B}) = \mathcal{D}\mathcal{A} + \mathcal{D}\mathcal{B} \quad \text{y} \quad \mathcal{D}(r\mathcal{A}) = r\mathcal{D}\mathcal{A}$$

En particular, si \mathcal{D}_c es la defusificación dada en (2.3), entonces

$$\mathcal{D}_c(\mathcal{A} + \mathcal{B}) = \mathcal{D}_c\mathcal{A} + \mathcal{D}_c\mathcal{B} \quad \text{and} \quad \mathcal{D}_c(r\mathcal{A}) = r\mathcal{D}_c\mathcal{A}.$$

DEMOSTRACIÓN: Es claro que:

$$\begin{aligned} \mathcal{D}(\mathcal{A} + \mathcal{B}) &= \sum_{i=1}^n \left(\beta_i \underline{\mathcal{A} + \mathcal{B}}_{\alpha_i} + \gamma_i \overline{\mathcal{A} + \mathcal{B}}_{\alpha_i} \right) = \sum_{i=1}^n \left(\beta_i (\underline{a}_{\alpha_i} + \underline{b}_{\alpha_i}) + \gamma_i (\bar{a}_{\alpha_i} + \bar{b}_{\alpha_i}) \right) \\ &= \sum_{i=1}^n (\beta_i \underline{a}_{\alpha_i} + \gamma_i \bar{a}_{\alpha_i}) + \sum_{i=1}^n (\beta_i \underline{b}_{\alpha_i} + \gamma_i \bar{b}_{\alpha_i}) = \mathcal{D}\mathcal{A} + \mathcal{D}\mathcal{B}. \end{aligned}$$

De manera análoga,

$$\begin{aligned} \mathcal{D}(r\mathcal{A}) &= \sum_{i=1}^n (\beta_i r \underline{\mathcal{A}}_{\alpha_i} + \gamma_i r \overline{\mathcal{A}}_{\alpha_i}) = \sum_{i=1}^n (\beta_i r \underline{\mathcal{A}}_{\alpha_i} + \gamma_i r \overline{\mathcal{A}}_{\alpha_i}) \\ &= r \sum_{i=1}^n (\beta_i \underline{a}_{\alpha_i} + \gamma_i \overline{a}_{\alpha_i}) = r \mathcal{D}\mathcal{A}. \end{aligned}$$

Como $\mathcal{D}_c\mathcal{A} = (1/2)\underline{a}_1 + (1/2)\overline{a}_1$ para todo $\mathcal{A} \in \mathcal{F}$ (es decir, $n = 1$ y $\beta_1 = \gamma_1 = 1/2$), las mismas propiedades se verifican para \mathcal{D}_c . ■

2.2.4. Variables aleatorias difusas

Las variables aleatorias difusas están cobrando un gran interés en la últimos años debido a su enorme aplicabilidad en estudios de tipo probabilístico. En primer lugar, recordamos el concepto de vector aleatorio.

Definición 2.2.14 (Vector aleatorio) *Sea $(\Omega, \mathfrak{A}, P)$ un espacio de probabilidad y sea $(\mathbb{R}^N, \mathfrak{B})$ su σ -álgebra de Borel. Un vector aleatorio es cualquier función medible $X : \Omega \rightarrow \mathbb{R}^N$. Si $N = 1$, diremos que X es una variable aleatoria.*

Las variables aleatorias reales clásicas se denominan variables aleatorias *crisp* (cuantitativas) en el contexto difuso. Existen otras nociones de *variable aleatoria difusa* que han sido introducidas por diferentes autores. Entre ellas, destacamos las siguientes dos:

- la definición debida a Kwäkernaak [35], que interpreta una variable aleatoria difusa como una percepción difusa de una variable aleatoria cuantitativa (crisp); y,
- la debida a Puri y Ralescu [48], que considera las variables aleatorias difusas como conjuntos aleatorios difusos.

En esta memoria, utilizamos la segunda concepción.

Definición 2.2.15 (Variable aleatoria difusa) *Sea $(\Omega, \mathfrak{A}, P)$ un espacio de probabilidad. Se dice que una función $\mathcal{X} : \Omega \rightarrow \mathcal{F}$ es una variable aleatoria difusa con valores en el espacio medible (Ω, \mathfrak{A}) si, para cada $\alpha \in [0, 1]$, el α -corte $\mathcal{X}_\alpha : \Omega \rightarrow \mathbb{R}$ dado por $X_\alpha(\omega) = X(\omega)_\alpha$ para todo $\omega \in \Omega$, es una variable aleatoria.*

Definición 2.2.16 (Variable aleatoria difusa trapezoidal) *Una función $\mathcal{X} : \Omega \rightarrow \mathcal{T}$ es una variable aleatoria difusa trapezoidal si la representación de \mathcal{X} , dada por*

$$(X^I/X^J/X^R/X^S) : \Omega \rightarrow \mathbb{R}^4,$$

es un vector aleatorio.

2.2.5. Valor esperado de una variable aleatoria difusa

En ocasiones es interesante resumir la información de una variable aleatoria difusa utilizando un único valor que nos permita conocer globalmente su comportamiento, realizar comparaciones, etc. La siguiente noción de valor esperado de una variable aleatoria difusa también se debe a Puri y Ralescu [48].

Definición 2.2.17 (Valor esperado de una variable aleatoria difusa trapezoidal) *El valor esperado de una variable aleatoria difusa trapezoidal \mathcal{X} , es el único conjunto difuso $E[\mathcal{X}]$ en \mathcal{T} cuya representación es*

$$(E[X^I]/E[X^J]/E[X^R]/E[X^S]).$$

CAPÍTULO 3

Medida de la distancia entre números difusos

Uno de los problemas con mayores aplicaciones en las diversas ramas de la ciencia es el de encontrar una noción de *distancia* entre objetos de diferentes clases que sea acorde al tipo de objetos que se manejan. Entre números reales, sin duda alguna, la distancia euclídea es la más conocida y utilizada. En un conjunto arbitrario X , una distancia es una función d que a cada par de elementos de X le asocia un número real, y que verifica cuatro archiconocidas propiedades (entre ellas, la *desigualdad triangular*). En el conjunto \mathcal{F} de los números difusos es relativamente sencillo inventarse una distancia real. Sin embargo, al tratarse \mathcal{F} de un conjunto de números difusos, no parece adecuado que la distancia entre dos números difusos sea un número real, debido esencialmente a la enorme pérdida de información que ello conlleva. Desde luego, parece más coherente que la distancia entre dos números difusos sea otro número difuso. En este campo no se han hecho grandes avances, especialmente cuando se afronta el conjunto \mathcal{F} en toda su extensión (entre ciertas clases de números difusos concretos, por ejemplo, entre números triangulares y/o trapezoidales, sí se pueden encontrar más aportaciones). La diferencia entre dos números difusos no sirve como dis-

tancia pues, entre otras cuestiones, no existe una forma adecuada de definir la diferencia entre dos números difusos que sea coherente con la suma, es decir, de manera que $\mathcal{A} - \mathcal{A}$ sea el número difuso $\tilde{0}$).

En el pasado, muchos/as autores/as han anunciado la presentación de diversas métricas en el conjunto \mathcal{F} , pero sus formas de medir la distancia entre dos números difusos no cumplían todas las condiciones para ser una auténtica métrica: en algunas ocasiones porque no cumplían la desigualdad triangular y, en otras, porque la distancia podía ser cero partiendo dos números difusos diferentes. En realidad, recientemente se habla más de *medida de similitud* (en inglés, *similarity measure*) que de *métrica* en el conjunto \mathcal{F} , pues se trata más de determinar cómo de parecidos o de diferentes son dos números difusos que de cumplir estrictamente con los axiomas de *distancia*. En este capítulo presentamos varias formas de medir la similitud entre dos números difusos, de manera que se obtenga un nuevo número difuso, de forma que, en muchos casos, nuestra definición puede considerarse una auténtica distancia. Nuestra metodología tiene en cuenta la forma geométrica de los números de partida y depende de algunas de las características esenciales de los mismos. Además, se podrá tener cierto control sobre las formas geométricas que se deseen obtener como distancias.

3.1. Medidas de distancias entre conjuntos arbitrarios

En esta sección vamos a introducir las características básicas que debe cumplir una aplicación que pretenda extender el concepto de *distancia* sobre un conjunto X , pero tomando valores sobre el propio conjunto X . Más adelante, aplicaremos esta visión al conjunto \mathcal{F} formado por todos los números difusos.

Antes de ello, recordemos que una *métrica* (o *distancia*) sobre un conjunto no vacío X es cualquier función $d : X \times X \rightarrow \mathbb{R}$ que verifique las siguientes

propiedades.

- *Positividad*: para cada $x, y \in X$ se tiene que $d(x, y) \geq 0$.
- *Identidad de los indiscernibles*: para cada $x, y \in X$ se tiene que $d(x, y) = 0$ si, y sólo si, $x = y$.
- *Simetría*: para cada $x, y \in X$ se tiene que $d(x, y) = d(y, x)$.
- *Desigualdad triangular*: para cada $x, y, z \in X$ se tiene que $d(x, z) \leq d(x, y) + d(y, z)$.

Si d es una métrica sobre X , se dice que el par (X, d) es un *espacio métrico*.

Con objeto de extender la definición anterior a un conjunto arbitrario X , debemos empezar por considerar una aplicación $d : X \times X \rightarrow X$ que tome valores sobre el propio conjunto X . Una de las primeras dificultades consiste en determinar el tipo de valores *no negativos* que conformarán el rango de la métrica. Para ello, utilizaremos un punto concreto del espacio y un orden parcial (o, en general, una relación binaria) sobre el conjunto. Además, este conjunto deberá estar provisto de una operación suma que sea coherente con el orden parcial de cara a verificar algún tipo de desigualdad triangular.

Planteado este problema sobre un conjunto X arbitrario, la respuesta no parece trivial. Por ejemplo, si X es la siguiente familia de etiquetas lingüísticas

$$X = \{ \text{“muy malo”, “malo”, “regular”, “bueno”, “muy bueno”} \},$$

entonces, ¿qué etiqueta puede considerarse como la distancia entre “regular” y “bueno”? Aún así, el conjunto anterior está intuitivamente ordenado. No obstante, si X es un conjunto finito de colores, ¿qué otro color puede considerarse como la distancia entre ellos?

Como una primera aproximación al problema antes mencionado, presentamos una forma natural de manejar una medida de la distancia entre dos puntos

de un conjunto X que toma valores sobre el mismo conjunto X . Un orden parcial y un punto fijo serán los conceptos clave que nos permiten considerar elementos *no negativos* sobre X . Más adelante presentaremos ejemplos de esta clase de distancias sobre el conjunto formado por todos los intervalos reales y compactos, y estudiaremos algunas propiedades.

Definición 3.1.1 *Una relación binaria sobre X es un subconjunto no vacío \mathcal{R} de $X \times X$. Por simplicidad, denotamos $x \trianglelefteq y$ si $(x, y) \in \mathcal{R}$, y diremos que \trianglelefteq es la relación binaria sobre X . Esta notación nos permite escribir $x \prec y$ cuando $x \trianglelefteq y$ y $x \neq y$. Escribiremos $y \trianglerighteq x$ cuando $x \trianglelefteq y$. Diremos que x e y son \trianglelefteq -comparables, y escribiremos $x \succsim y$, si $x \trianglelefteq y$ o $y \trianglelefteq x$. Una relación binaria \trianglelefteq sobre X es reflexiva si $x \trianglelefteq x$ para cada $x \in X$; es transitiva si $x \trianglelefteq z$ para cada $x, y, z \in X$ tales que $x \trianglelefteq y$ e $y \trianglelefteq z$; y es antisimétrica si $x \trianglelefteq y$ e $y \trianglelefteq x$ implican que $x = y$.*

Una relación reflexiva y transitiva \preceq sobre X es un preorden (o un quasior-den) sobre X . En tal caso, (X, \preceq) es un espacio preordenado. Si un preorden \preceq es también antisimétrico, entonces \preceq se llama un orden parcial, y (X, \preceq) es un espacio parcialmente ordenado (o un conjunto parcialmente ordenado). Utilizaremos el símbolo \trianglelefteq para denotar una relación binaria general sobre X , y el símbolo \preceq para una relación binaria reflexiva sobre X (por ejemplo, un preorden o un orden parcial).

Definición 3.1.2 *Sea 0_X un punto de un conjunto X dotado de una relación binaria \preceq (usualmente, supondremos que se trata de un orden parcial). Consideremos el conjunto*

$$X_{0, \preceq}^+ = \{x \in X : 0_X \preceq x\}$$

y sea $s : X_{0, \preceq}^+ \times X_{0, \preceq}^+ \rightarrow X_{0, \preceq}^+$ una aplicación. Una función de distancia sobre $(X, 0_X, \preceq, s)$ (o una métrica) es una aplicación $d : X \times X \rightarrow X_{0, \preceq}^+$ que verifica,

para cualquier $x, y, z \in X$,

- (i) $d(x, x) = 0_X$,
- (ii) si $d(x, y) = 0_X$, entonces $x = y$;
- (iii) $d(x, y) = d(y, x)$,
- (iv) $d(x, z) \preceq s(d(x, y), d(y, z))$

En tal caso, diremos que $(X, 0_X, s)$ es un espacio métrico con respecto al orden parcial \preceq . La función d es:

- una pseudométrica si verifica (i), (iii) y (iv);
- una semimétrica (sobre $(X, 0_X, \preceq)$) si verifica (i), (ii) y (iii);
- una pseudosemimétrica (sobre $(X, 0_X, \preceq)$) si verifica (i) y (iii).

Nótese que la positividad de la métrica d se impone al tomar ésta valores sobre el subconjunto $X_{0, \preceq}^+$, y la identidad de los indiscernibles se desdobra en las propiedades (i) y (ii). Obsérvese también que, aunque una semimétrica o una pseudosemimétrica no están obligadas a cumplir la propiedad (iv), sí es verdad que deben estar valuadas en el conjunto $X_{0, \preceq}^+$, donde juega un papel fundamental la relación binaria \preceq .

Ejemplo 3.1.3 Cuando X coincide con \mathbb{R} , \mathbb{R}_0^+ o un intervalo del tipo $[0, r]$ (donde r es un número positivo real), provisto del orden parcial usual \leq y la suma de números, la noción anterior de métrica coincide con el concepto de distancia en sentido clásico.

Ejemplo 3.1.4 Un conjunto parcialmente ordenado con dos puntos diferentes pero comparables puede ser siempre provisto de una métrica en el sentido de la Definición 3.1.2. Por ejemplo, sean 0_X y 1_X dos puntos diferentes de un conjunto X parcialmente ordenado tal que $0_X \preceq 1_X$, y definamos $d(x, y) = 0_X$,

si $x = y$, y $d(x, y) = 1_X$, si $x \neq y$. Entonces d es una semimétrica sobre $(X, 0_X, \preceq)$. Además, si $s : X_{0, \preceq}^+ \times X_{0, \preceq}^+ \rightarrow X_{0, \preceq}^+$ es una función simétrica tal que $1_X \preceq s(1_X, x)$ para todo $x \in X_{0, \preceq}^+$, entonces d es una función de distancia sobre $(X, 0_X, \preceq, s)$. Tiene sentido que esta métrica se denomine métrica discreta sobre X .

Una de las equivocaciones más frecuentes cuando se aplica el método de mínimos cuadrados consiste en decir que la aplicación $\mathbf{d} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_0^+$ dada por

$$\mathbf{d}(x, y) = (x - y)^2 \quad \text{para cada } x, y \in \mathbb{R},$$

es una métrica real. Sin embargo, esto no es cierto, ya que no cumple la desigualdad triangular. Por ejemplo, nótese que

$$(2 - 1)^2 > (2 - 1.5)^2 + (1.5 - 1)^2.$$

En este caso, debemos hablar de una semimétrica, como demostramos en el siguiente enunciado.

Lema 3.1.5 Si $p, q > 0$, entonces la función $\phi : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_0^+$, dada por

$$\phi(x, y) = q |x - y|^p \quad \text{para cada } x, y \in \mathbb{R},$$

es una semimétrica sobre $(\mathbb{R}, 0, \leq)$.

Y si $0 < p \leq 1$, entonces ϕ es una métrica sobre $(\mathbb{R}, 0, \leq, +)$.

DEMOSTRACIÓN: Es claro que $\phi(x, y) = 0$ si, y sólo si, $x = y$, y que $\phi(x, y) = \phi(y, x)$. Si $p = 1$, entonces $q |x - z| \leq q |x - y| + q |y - z|$, lo que significa que ϕ es una métrica. Supongamos que $0 < p < 1$ y vamos a demostrar que

$$|x - z|^p \leq |x - y|^p + |y - z|^p. \quad (3.1)$$

Fijemos $x, z \in \mathbb{R}$. Si $x = z$, entonces (3.1) es trivial. Supongamos que $x < z$ sin pérdida de generalidad. Consideremos la función $h : \mathbb{R} \rightarrow \mathbb{R}$ dada por $h(y) =$

$|x - y|^p + |y - z|^p - |x - z|^p$ para cada $y \in \mathbb{R}$. Claramente, h es continua sobre \mathbb{R} y $h(x) = h(z) = 0$. Llamemos $a = (z - x)^p$. Entonces, podemos descomponer h a trozos de la siguiente forma:

$$h(y) = \begin{cases} (x - y)^p + (z - y)^p - a, & \text{si } y < x, \\ (y - x)^p + (z - y)^p - a, & \text{si } x \leq y \leq z, \\ (y - x)^p + (y - z)^p - a, & \text{si } y > z. \end{cases}$$

No es difícil demostrar que $h'(y) < 0$ para cada $y < x$ y que $h'(y) > 0$ para cada $y > z$. Por tanto, h es estrictamente decreciente en $]-\infty, x[$ y estrictamente creciente en $]z, \infty[$. De esta forma, h posee, al menos, un mínimo absoluto, que se encuentra en el intervalo $[x, z]$. No obstante, como $h''(y) < 0$ para cada $y \in]x, z[$, deducimos que su mínimo no está en el intervalo $]x, z[$, por lo que, obligatoriamente, debe estar en el conjunto $\{x, z\}$. Dado que $h(x) = h(z) = 0$, concluimos que $h(y) \geq 0$ para cada $y \in \mathbb{R}$, por lo que la desigualdad (3.1) es cierta. ■

Repitiendo los argumentos de la demostración anterior y teniendo en cuenta que la función $x \mapsto x^k$ es estrictamente creciente en \mathbb{R}_0^+ cuando $k > 0$, se prueba la siguiente extensión.

Lema 3.1.6 Si $p, q, k > 0$, entonces la función $\psi : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$, dada por

$$\psi(x, y) = q |x^k - y^k|^p \quad \text{para cada } x, y \in \mathbb{R},$$

es una semimétrica sobre $(\mathbb{R}_0^+, 0, \leq)$.

Y si $0 < p \leq 1$, entonces ψ es una métrica sobre $(\mathbb{R}_0^+, 0, \leq, +)$.

3.2. Medidas de distancias entre intervalos compactos

Desde nuestro punto de vista, existe una doble motivación para considerar distancias entre intervalos reales compactos. Por un lado, los α -cortes de los números difusos que consideramos en esta memoria son intervalos compactos que sirven para caracterizar completamente a los números difusos de los que proceden. Por otro lado, las operaciones básicas entre números difusos se definen en términos de sus α -cortes. Por esta razón, dado que las métricas que presentaremos entre números difusos estarán expresadas en términos de dichos conjuntos de nivel, vamos a introducir algunas distancias entre intervalos compactos en el sentido de la Definición 3.1.2.

En lo que sigue, sea \mathcal{I} la familia formada por todos los subintervalos compactos no vacíos de \mathbb{R} , y sea $0_{\mathcal{I}}$ el intervalo $0_{\mathcal{I}} = [0, 0]$. La suma de intervalos de \mathcal{I} viene dada por:

$$s_{\mathcal{I}}([a, b], [c, e]) = [a, b] + [c, e] = [a + c, b + e] \quad (3.2)$$

para cada $[a, b], [c, e] \in \mathcal{I}$. Veamos una primera forma de aplicar las nociones anteriores.

Ejemplo 3.2.1 *Consideremos sobre \mathcal{I} el orden parcial \sqsubseteq dado por*

$$[a, b] \sqsubseteq [c, e] \quad \text{si } a \leq c \text{ y } b \leq e.$$

En este contexto, el conjunto

$$\mathcal{I}_{0_{\mathcal{I}}, \sqsubseteq}^+ = \{[a, b] \in \mathcal{I} : [0, 0] \sqsubseteq [a, b]\} = \{[a, b] \in \mathcal{I} : 0 \leq a \leq b\}$$

coincide con la familia de intervalos compactos de extremos no negativos. Es conocido (véase [1]) que la función $d : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_{0_{\mathcal{I}}, \sqsubseteq}^+$, dada por

$$d([a, b], [c, e]) = [\min(|a - c|, |b - e|), \max(|a - c|, |b - e|)] \quad (3.3)$$

para todo $[a, b], [c, e] \in \mathcal{I}$, es una distancia sobre $(\mathcal{I}, 0_{\mathcal{I}}, \subseteq, s_{\mathcal{I}})$ en el sentido de la Definición 3.1.2.

Aunque el ejemplo anterior cumple todas las características de una métrica sobre el conjunto \mathcal{I} en el sentido de la Definición 3.1.2, para nuestros propósitos, utilizaremos una métrica diferente en este conjunto. Para describirla, observemos que cualquier aplicación arbitraria $d : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}$ está determinada por la forma de asociarle a cada pareja de intervalos compactos otro intervalo caracterizado por sus extremos inferior y superior. De esta forma, denotaremos $d = [\underline{d}, \bar{d}]$, donde $\underline{d}, \bar{d} : \mathcal{I} \times \mathcal{I} \rightarrow \mathbb{R}$ miden los extremos inferior y superior del intervalo, es decir,

$$d(I, J) = [\underline{d}(I, J), \bar{d}(I, J)] \quad \text{para cada } I, J \in \mathcal{I}.$$

Sobre el conjunto \mathcal{I} , consideremos el orden parcial \subseteq dado por la inclusión natural de subconjuntos, es decir,

$$[a, b] \subseteq [c, e] \quad \text{si} \quad c \leq a \leq b \leq e.$$

Definimos $\delta : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}$ como

$$\delta([a, b], [c, e]) = [-|a - c|, |b - e|] \tag{3.4}$$

para todo $[a, b], [c, e] \in \mathcal{I}$, o sea,

$$\underline{\delta}([a, b], [c, e]) = -|a - c| \quad \text{y} \quad \bar{\delta}([a, b], [c, e]) = |b - e|.$$

Lema 3.2.2 Para cada intervalos $I, I', J, J', K \in \mathcal{I}$, se verifican las siguientes propiedades.

1. $\delta : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_{0_{\mathcal{I}}, \subseteq}^+$ es una métrica sobre $(\mathcal{I}, 0_{\mathcal{I}}, \subseteq, s_{\mathcal{I}})$.

2. Si $I \subseteq I'$ y $J \subseteq J'$, entonces $\delta(I, I') + \delta(J, J') = \delta(I + J, I' + J')$.
3. $\delta(I + K, J + K) = \delta(I, J)$.
4. Si $I \subseteq J \cap K$ y $\delta(I, J) = \delta(I, K)$, entonces $J = K$.
5. Si $I \subseteq J$, entonces $I + K \subseteq J + K$.

DEMOSTRACIÓN: Supongamos que $I = [a, b]$, $I' = [a', b']$, $J = [c, e]$, $J' = [c', e']$ y $K = [f, g]$.

(1) En primer lugar, dado que $-|a - c| \leq 0 \leq |b - e|$, es claro que $0_{\mathcal{I}} \subseteq \delta(I, J) = [-|a - c|, |b - e|]$. Así, $\delta(I, J) \in \mathcal{I}_{0_{\mathcal{I}}, \subseteq}^+$ para cada $I, J \in \mathcal{I}$. Por otro lado, $\delta(I, I) = [0, 0] = 0_{\mathcal{I}}$. Además, $\delta(I, J) = [-|a - c|, |b - e|] = 0_{\mathcal{I}}$ si, y sólo si, $a = c$ y $b = e$, en cuyo caso $I = J$. También es obvia la simetría de δ . Finalmente, dado que

$$|b - g| \leq |b - e| + |e - g| \quad \text{y} \quad -|a - f| \geq -|a - c| - |c - f|,$$

deducimos que

$$\begin{aligned} \delta(I, K) &= [-|a - f|, |b - g|] \subseteq [-|a - c| - |c - f|, |b - e| + |e - g|] \\ &= [-|a - c|, |b - e|] + [-|c - f|, |e - g|] \\ &= \delta(I, J) + \delta(J, K) = s_{\mathcal{I}}(\delta(I, J), \delta(J, K)), \end{aligned}$$

por lo que también se cumple la desigualdad triangular, y $\delta : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_{0_{\mathcal{I}}, \subseteq}^+$ es una métrica sobre $(\mathcal{I}, 0_{\mathcal{I}}, \subseteq, s_{\mathcal{I}})$.

- (2) Si $I \subseteq I'$ y $J \subseteq J'$, entonces $a' \leq a \leq b \leq b'$ y $c' \leq c \leq e \leq e'$. De esta

forma,

$$\begin{aligned}
\delta(I + J, I' + J') &= \delta([a, b] + [c, e], [a', b'] + [c', e']) \\
&= \delta([a + c, b + e], [a' + c', b' + e']) \\
&= [-|(a + c) - (a' + c')|, |(b + e) - (b' + e')|] \\
&= [-|(a - a') + (c - c')|, |(b - b') + (e - e')|] \\
&= [-((a - a') + (c - c')), (b' - b) + (e' - e)] \\
&= [-(a - a'), b' - b] + [-(c - c'), e' - e] \\
&= [-|a - a'|, |b' - b|] + [-|c - c'|, |e' - e|] \\
&= \delta(I, I') + \delta(J, J').
\end{aligned}$$

(3) Se deduce de que

$$\begin{aligned}
\delta(I + K, J + K) &= \delta([a, b] + [f, g], [c, e] + [f, g]) \\
&= \delta([a + f, b + g], [c + f, e + g]) \\
&= [-|(a + f) - (c + f)|, |(b + g) - (e + g)|] \\
&= [-|a - c|, |b - e|] = \delta(I, J).
\end{aligned}$$

(4) Supongamos que $I \subseteq J \cap K$ (es decir, $c \leq a \leq b \leq e$ y $f \leq a \leq b \leq g$) y que $\delta(I, J) = \delta(I, K)$. Esto significa que

$$\begin{aligned}
[c - a, e - b] &= [-|a - c|, |b - e|] = \delta(I, J) = \delta(I, K) \\
&= [-|a - f|, |b - g|] = [f - a, g - b].
\end{aligned}$$

Así, $c - a = f - a$ y $e - b = g - b$. Por tanto, $[a, b] = [c, e]$, es decir, $J = K$.

(5) Supongamos que $I \subseteq J$, es decir, $c \leq a \leq b \leq e$. Entonces $c + f \leq a + f$ y $b + g \leq e + g$. Como $a \leq b$ y $f \leq g$, entonces $a + f \leq b + g$. Uniendo todas las desigualdades, deducimos que $c + f \leq a + f \leq b + g \leq e + g$, lo que significa que

$$I + K = [a, b] + [f, g] = [a + f, b + g] \subseteq [c + f, e + g] = [c, e] + [f, g] = J + K,$$

lo que concluye la demostración. ■

La diferencia entre la aplicación d definida en (3.3) y la aplicación δ dada en (3.4) se observa, por ejemplo, sobre los siguientes intervalos, pues:

$$d([3, 5], [-1, 6]) = [1, 4] \quad \text{and} \quad \delta([3, 5], [-1, 6]) = [-4, 1].$$

Cuando la suma de intervalos compactos y la métrica δ se aplican a ciertos conjuntos de nivel de números difusos, encontramos las siguientes relaciones con la suma de números difusos.

Lema 3.2.3 Si $\mathcal{A}, \mathcal{B} \in \mathcal{F}$, entonces

$$\begin{aligned} \ker(\mathcal{A} + \mathcal{B}) &= \ker \mathcal{A} + \ker \mathcal{B}, \\ \text{supp}(\mathcal{A} + \mathcal{B}) &= \text{supp} \mathcal{A} + \text{supp} \mathcal{B}, \\ \text{spr}(\mathcal{A} + \mathcal{B}) &= \text{spr} \mathcal{A} + \text{spr} \mathcal{B} \quad \text{y} \\ \delta(\ker \mathcal{A}, \text{supp} \mathcal{A}) &= [\underline{a}_0 - \underline{a}_1, \bar{a}_0 - \bar{a}_1]. \end{aligned}$$

DEMOSTRACIÓN: Por definición,

$$\begin{aligned} \ker(\mathcal{A} + \mathcal{B}) &= [\underline{\mathcal{A}} + \underline{\mathcal{B}}_1, \overline{\mathcal{A}} + \overline{\mathcal{B}}_1] = [\underline{a}_1 + \underline{b}_1, \bar{a}_1 + \bar{b}_1] \\ &= [\underline{a}_1, \bar{a}_1] + [\underline{b}_1, \bar{b}_1] = \ker \mathcal{A} + \ker \mathcal{B}. \end{aligned}$$

De forma similar,

$$\begin{aligned} \text{supp}(\mathcal{A} + \mathcal{B}) &= [\underline{\mathcal{A}} + \underline{\mathcal{B}}_0, \overline{\mathcal{A}} + \overline{\mathcal{B}}_0] = [\underline{a}_0 + \underline{b}_0, \bar{a}_0 + \bar{b}_0] \\ &= [\underline{a}_0, \bar{a}_0] + [\underline{b}_0, \bar{b}_0] = \text{supp} \mathcal{A} + \text{supp} \mathcal{B}. \end{aligned}$$

Análogamente,

$$\begin{aligned} \text{spr}(\mathcal{A} + \mathcal{B}) &= (\overline{\mathcal{A}} + \overline{\mathcal{B}}_1 + \underline{\mathcal{A}} + \underline{\mathcal{B}}_1)/2 = (\bar{a}_1 + \bar{b}_1 - \underline{a}_1 - \underline{b}_1)/2 \\ &= (\bar{a}_1 - \underline{a}_1)/2 + (\bar{b}_1 - \underline{b}_1)/2 = \text{spr} \mathcal{A} + \text{spr} \mathcal{B}. \end{aligned}$$

La última propiedad se deduce de que

$$\begin{aligned} \delta(\ker \mathcal{A}, \text{supp} \mathcal{A}) &= \delta([\underline{a}_1, \bar{a}_1], [\underline{a}_0, \bar{a}_0]) = [-|\underline{a}_1 - \underline{a}_0|, |\bar{a}_1 - \bar{a}_0|] \\ &= [\underline{a}_0 - \underline{a}_1, \bar{a}_0 - \bar{a}_1]. \end{aligned}$$

■

3.3. Medidas de distancia entre números difusos

En esta sección, introducimos una familia de medidas de distancia entre números difusos arbitrarios que dependen de una defusificación, de una distancia intervalar en \mathcal{I} con respecto a un orden parcial \sqsubseteq sobre \mathcal{I} , y algunas pseudosemimétricas reales. Antes de esto, describimos una relación binaria sobre el conjunto \mathcal{F} que, particularizada a la familia de todos los números difusos trapezoidales, se convierte en un orden parcial.

A lo largo de esta sección, sea $(\mathcal{D}, d, \sqsubseteq)$ una terna donde $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$ es una defusificación, \sqsubseteq es un orden parcial sobre \mathcal{I} , y $d = [d, \bar{d}] : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_{0_{\mathcal{I}}, \sqsubseteq}^+$ es una distancia intervalar sobre $(\mathcal{I}, 0_{\mathcal{I}}, \sqsubseteq, s_{\mathcal{I}})$ en el sentido de la Definición 3.1.2.

3.3.1. Un orden parcial sobre los números difusos trapezoidales

Definición 3.3.1 *Para cualesquiera números difusos $\mathcal{A}, \mathcal{B} \in \mathcal{F}$, escribiremos $\mathcal{A} \preceq \mathcal{B}$ con respecto a $(\mathcal{D}, d, \sqsubseteq)$ si $\mathcal{D}\mathcal{A} \leq \mathcal{D}\mathcal{B}$, $\text{spr } \mathcal{A} \leq \text{spr } \mathcal{B}$ y $d(\ker \mathcal{A}, \text{supp } \mathcal{A}) \sqsubseteq d(\ker \mathcal{B}, \text{supp } \mathcal{B})$. Si consideramos $(\mathcal{D}_c, \delta, \sqsubseteq)$, sólo escribiremos $\mathcal{A} \preceq \mathcal{B}$.*

En la definición anterior, podríamos considerar distancias entre diferentes α -cortes de los números difusos \mathcal{A} y \mathcal{B} . Sin embargo, principalmente nos interesan los números difusos trapezoidales porque tienen la ventaja de ser representaciones simples que puede medirse e implementarse con facilidad por investigadores/as de diferentes ámbitos. Por lo tanto, sólo hemos considerado sus núcleos y sus soportes ya que son los conjuntos de nivel que caracterizan dichos números difusos.

Teorema 3.3.2 **La relación \preceq con respecto a $(\mathcal{D}, d, \sqsubseteq)$ verifica las siguientes**

tes propiedades para todos los números difusos $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathcal{F}$.

1. Es reflexiva y transitiva.
2. $\tilde{0} \preceq \mathcal{A}$ con respecto a $(\mathcal{D}, d, \sqsubseteq)$ si, y sólo si, $\mathcal{D}\tilde{0} \leq \mathcal{D}\mathcal{A}$.
3. Supongamos que $\mathcal{D}\mathcal{C} = \alpha \underline{c}_1 + \beta \bar{c}_1$ para cada $\mathcal{C} \in \mathcal{T}$ (donde $\alpha, \beta \in \mathbb{R}$ y $\alpha + \beta \neq 0$). Si \mathcal{A} y \mathcal{B} son números difusos trapezoidales, $\mathcal{A} \preceq \mathcal{B}$ con respecto a $(\mathcal{D}, \delta, \sqsubseteq)$ y $\mathcal{B} \preceq \mathcal{A}$ con respecto a $(\mathcal{D}, \delta, \sqsubseteq)$, entonces $\mathcal{A} = \mathcal{B}$.

En particular, \preceq es un orden parcial sobre \mathcal{T} . Además, si $\mathcal{A} \preceq \mathcal{B}$, entonces $\mathcal{A} + \mathcal{C} \preceq \mathcal{B} + \mathcal{C}$ y $r\mathcal{A} \preceq r\mathcal{B}$ para todo $r > 0$.

DEMOSTRACIÓN: El primer apartado es trivial. Para (2), observemos que $\text{spr } \tilde{0} = 0 \leq \text{spr } \mathcal{A}$ y $d(\ker \tilde{0}, \text{supp } \tilde{0}) = d(0_I, 0_I) = 0_I \sqsubseteq d(\ker \mathcal{A}, \text{supp } \mathcal{A})$. Con respecto al tercer apartado, si $\mathcal{A} \preceq \mathcal{B}$ y $\mathcal{B} \preceq \mathcal{A}$, entonces

$$\left\{ \begin{array}{l} \mathcal{D}\mathcal{A} = \mathcal{D}\mathcal{B}, \\ \text{spr } \mathcal{A} = \text{spr } \mathcal{B}, \\ \delta(\ker \mathcal{A}, \text{supp } \mathcal{A}) = \\ = \delta(\ker \mathcal{B}, \text{supp } \mathcal{B}) \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \alpha \underline{a}_1 + \beta \bar{a}_1 = \alpha \underline{b}_1 + \beta \bar{b}_1, \\ \bar{a}_1 - \underline{a}_1 = \bar{b}_1 - \underline{b}_1, \\ \underline{a}_1 - \underline{a}_0 = \underline{b}_1 - \underline{b}_0, \\ \bar{a}_0 - \bar{a}_1 = \bar{b}_0 - \bar{b}_1 \end{array} \right\} \Leftrightarrow$$

$$\Leftrightarrow \left\{ \begin{array}{l} \underline{a}_1 = \underline{b}_1, \\ \bar{a}_1 = \bar{b}_1, \\ \underline{a}_0 = \underline{b}_0, \\ \bar{a}_0 = \bar{b}_0 \end{array} \right\} \Leftrightarrow \left\{ \begin{array}{l} \ker \mathcal{A} = \ker \mathcal{B}, \\ \text{supp } \mathcal{A} = \text{supp } \mathcal{B} \end{array} \right\}.$$

Como el núcleo y el soporte caracterizan completamente a un número difuso trapezoidal, entonces $\mathcal{A} = \mathcal{B}$.

Ahora supongamos que $\mathcal{A} \preceq \mathcal{B}$. Utilizando el Lema 3.2.3,

$$\mathcal{D}_c(\mathcal{A} + \mathcal{C}) = \mathcal{D}_c\mathcal{A} + \mathcal{D}_c\mathcal{C} \leq \mathcal{D}_c\mathcal{B} + \mathcal{D}_c\mathcal{C} = \mathcal{D}_c(\mathcal{B} + \mathcal{C}).$$

Como $\text{spr } \mathcal{A} \sqsubseteq \text{spr } \mathcal{B}$, el Lema 3.2.3 y el apartado (5) del Lema 5 nos muestra que

$$\text{spr}(\mathcal{A} + \mathcal{C}) = \text{spr } \mathcal{A} + \text{spr } \mathcal{C} \leq \text{spr } \mathcal{B} + \text{spr } \mathcal{C} = \text{spr}(\mathcal{B} + \mathcal{C}).$$

Finalmente, como

$$\delta(\ker \mathcal{A}, \text{supp } \mathcal{A}) \subseteq \delta(\ker \mathcal{B}, \text{supp } \mathcal{B}),$$

por el apartado (5) del Lema 3.2.2, tenemos que

$$\delta(\ker \mathcal{A}, \text{supp } \mathcal{A}) + \delta(\ker \mathcal{C}, \text{supp } \mathcal{C}) \subseteq \delta(\ker \mathcal{B}, \text{supp } \mathcal{B}) + \delta(\ker \mathcal{C}, \text{supp } \mathcal{C}),$$

y el apartado (2) del Lema 3.2.2 implica que

$$\delta(\ker \mathcal{A} + \ker \mathcal{C}, \text{supp } \mathcal{A} + \text{supp } \mathcal{C}) \subseteq \delta(\ker \mathcal{B} + \ker \mathcal{C}, \text{supp } \mathcal{B} + \text{supp } \mathcal{C}),$$

es decir,

$$\delta(\ker(\mathcal{A} + \mathcal{C}), \text{supp}(\mathcal{A} + \mathcal{C})) \subseteq \delta(\ker(\mathcal{B} + \mathcal{C}), \text{supp}(\mathcal{B} + \mathcal{C})).$$

Por lo tanto,

$$\mathcal{A} + \mathcal{C} \preceq \mathcal{B} + \mathcal{C}.$$

La misma demostración es válida para probar que $r\mathcal{A} \preceq r\mathcal{B}$. ■

3.3.2. Medidas de similitud entre números difusos

No es difícil considerar una distancia sobre \mathcal{T} . Por ejemplo, sean $\phi_1, \phi_2, \phi_3, \phi_4 : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_0^+$ cuatro métricas sobre \mathbb{R} y definamos

$$D((a/b/c/e), (a'/b'/c'/e')) = \sim (\phi_1(a, a') + \phi_2(b, b') + \phi_3(c, c') + \phi_4(e, e')),$$

donde $\sim r = \tilde{r}$ denota la forma canónica de ver un número real r como un número difuso \tilde{r} . En tal caso, D es una distancia sobre \mathcal{T} . Sin embargo, estas métricas son esencialmente reales, por lo que conllevan una pérdida de información muy importante. Además, no es fácil generalizarlas a la familia de todos los números difusos. Por ello, estas métricas no son demasiado interesantes en el ambiente difuso.

En esta subsección describimos una aplicación $D : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{F}$ que asigna, a cada par de números difusos \mathcal{A} y \mathcal{B} , otro número difuso $D(\mathcal{A}, \mathcal{B})$ que trata de dar una idea de lo parecidos o de lo diferentes que son los números difusos \mathcal{A} y \mathcal{B} . La definición que vamos a utilizar se basa en las siguientes consideraciones geométricas.

- El centro de $D(\mathcal{A}, \mathcal{B})$ debe depender sólo de los puntos que el/la investigador/a considere que son más importantes para \mathcal{A} y \mathcal{B} , es decir, el/la investigador/a debe considerar una defusificación $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$ que asocie un punto concreto de cada número difuso (su media, su ambigüedad, etc.). La distancia entre estos puntos será determinante para calcular $D(\mathcal{A}, \mathcal{B})$. Como consecuencia, si ϕ_1 mide distancias entre números reales, utilizaremos la propiedad

$$\mathcal{D}_c(D(\mathcal{A}, \mathcal{B})) = \phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B})$$

para construir $D(\mathcal{A}, \mathcal{B})$.

- La amplitud de $D(\mathcal{A}, \mathcal{B})$ debe depender, esencialmente, de las amplitudes de \mathcal{A} y \mathcal{B} . En este sentido, cuanto mayores sean los radios de \mathcal{A} y de \mathcal{B} , mayor debería ser el radio de su medida de similitud. Como consecuencia, entendemos que debe cumplirse que

$$\text{spr } D(\mathcal{A}, \mathcal{B}) = \psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B})$$

donde ψ_2 mide la distancia entre números reales no negativos.

- El/la investigador/a debe tener el control sobre el tipo de números difusos que desea obtener como medidas de distancia. De esta forma, introduciremos ciertas funciones h_1 y h_2 que nos permitirán decidir si la distancia obtenida es un número difuso triangular, trapezoidal, rectangular, LR o crisp, dependiendo de los objetivos del estudio.
- En el soporte de $D(\mathcal{A}, \mathcal{B})$ tendrán influencia los núcleos y los soportes de \mathcal{A} y \mathcal{B} , y la forma en la que medimos la distancia entre ellos.

- Si el/la investigador/a lo considera importante, la forma en la que se determina la distancia entre dos números difusos deberá ser capaz de extender la notación real en dos sentidos: (1) Si a y b son números reales, entonces $D(\tilde{a}, \tilde{b})$ debe de ser coherente con la forma en la que el investigador mide la distancia real entre a y b (que, por ejemplo, podría ser $|a - b|$); (2) algunos/as investigadores/as han utilizado, en la familia formada por todos los números difusos triangulares, medidas de distancia reales como la siguiente:

$$D[(a/b/c), (a'/b'/c')] = (a - a')^2 + (b - b')^2 + (c - c')^2. \quad (3.5)$$

Como se puede comprobar fácilmente, esta medida de distancia no es una distancia verdadera en el sentido clásico porque no verifica la desigualdad triangular. Sin embargo, tiene una propiedades muy buenas (por ejemplo, en regresión difusa, es considerada como función pérdida en el método de mínimos cuadrados) que deberán poder contemplarse en las medidas que presentemos. Como la expresión (3.5) define una pseudosemimétrica en la familia de todos los números difusos triangulares, también vamos a considerar esta familia de medidas de similitud.

Como consecuencia de las consideraciones previas, proponemos la siguiente familia de medidas de similitud entre los números difusos. Observemos que, en algunos casos particulares, una métrica parecida fue considerada en [61].

Teorema 3.3.3 Sea $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$ una defusificación, sean $q_1, q_2 \geq 0$ dos constantes no negativas, sean $h_1, h_2 : \mathbb{I} \rightarrow [0, \infty)$ dos funciones no crecientes y continuas a la izquierda en \mathbb{I} , sean $\phi_1, \phi_3, \phi_4 : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_0^+$ tres pseudosemimétricas sobre \mathbb{R} y sea $\psi_2 : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ una pseudosemimétrica sobre \mathbb{R}_0^+ . Entonces, para cualesquiera números difusos $\mathcal{A}, \mathcal{B} \in \mathcal{F}$, existe un único

número difuso $\mathcal{C} \in \mathcal{F}$ tal que, para todo $\alpha \in \mathbb{I}$,

$$\begin{aligned} \underline{\mathcal{C}}_\alpha &= q_1\phi_1(\mathcal{DA}, \mathcal{DB}) - q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) \\ &\quad - h_1(\alpha) \phi_3(\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})); \end{aligned} \quad (3.6)$$

$$\begin{aligned} \overline{\mathcal{C}}_\alpha &= q_1\phi_1(\mathcal{DA}, \mathcal{DB}) + q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) \\ &\quad + h_2(\alpha) \phi_4(\overline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \overline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})). \end{aligned} \quad (3.7)$$

DEMOSTRACIÓN: Como $\phi_1, \phi_3, \phi_4, \psi_2, h_1, h_2 \geq 0$ y $q_1, q_2 \geq 0$, entonces, para cada $\alpha \in \mathbb{I}$, $\underline{\mathcal{C}}_\alpha$ y $\overline{\mathcal{C}}_\alpha$ verifican

$$\underline{\mathcal{C}}_\alpha \leq q_1\phi_1(\mathcal{DA}, \mathcal{DB}) \leq \overline{\mathcal{C}}_\alpha.$$

De esta forma, $[\underline{\mathcal{C}}_\alpha, \overline{\mathcal{C}}_\alpha]$ es un subintervalo compacto de \mathbb{R} . Como $h_1, h_2 : \mathbb{I} \rightarrow [0, \infty)$ son dos funciones no crecientes y continuas a la izquierda sobre \mathbb{I} , deducimos que $\underline{\mathcal{C}}$ y $\overline{\mathcal{C}}$ son funciones continuas por la izquierda sobre \mathbb{I} , $\underline{\mathcal{C}}$ es no decreciente y $\overline{\mathcal{C}}$ es no creciente. Utilizando el Lema 2.2.6, estas funciones determinan un único número difuso $\mathcal{C} \in \mathcal{F}$ cuyos α -cortes vienen dados por las expresiones (3.6)-(3.7). ■

En la siguiente definición presentamos una forma de asociarle a cada par de números difusos \mathcal{A} y \mathcal{B} otro número difuso $D(\mathcal{A}, \mathcal{B})$ que nos servirá, en lo sucesivo, para determinar la semejanza o la diferencia entre los números \mathcal{A} y \mathcal{B} y, en algunos casos, servirá como métrica en algunos subconjuntos de \mathcal{F} .

Definición 3.3.4 Sea $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$ una defusificación, sean $q_1, q_2 \geq 0$ dos constantes no negativas, sean $h_1, h_2 : \mathbb{I} \rightarrow [0, \infty)$ dos funciones no crecientes y continuas por la izquierda sobre \mathbb{I} , sean $\phi_1, \phi_3, \phi_4 : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_0^+$ tres pseudo-semimétricas sobre \mathbb{R} y sea $\psi_2 : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ una pseudosemimétrica sobre \mathbb{R}_0^+ . Para cualesquiera números difusos $\mathcal{A}, \mathcal{B} \in \mathcal{F}$, denotaremos por $D(\mathcal{A}, \mathcal{B})$ al único número difuso \mathcal{C} dado en el Teorema 3.3.3, es decir, para cada $\alpha \in \mathbb{I}$, los

extremos del α -corte de $D(\mathcal{A}, \mathcal{B})$ están dados por las siguientes expresiones:

$$\begin{aligned} \underline{D(\mathcal{A}, \mathcal{B})}_\alpha &= q_1 \phi_1(\mathcal{DA}, \mathcal{DB}) - q_2 \psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) \\ &\quad - h_1(\alpha) \phi_3(\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})); \end{aligned} \quad (3.8)$$

$$\begin{aligned} \overline{D(\mathcal{A}, \mathcal{B})}_\alpha &= q_1 \phi_1(\mathcal{DA}, \mathcal{DB}) + q_2 \psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) \\ &\quad + h_2(\alpha) \phi_4(\overline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \overline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})). \end{aligned} \quad (3.9)$$

La definición anterior nos permite considerar una aplicación $D : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{F}$ que depende de las constantes no negativas q_1 y q_2 , de la defusificación $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$, de las funciones h_1 y h_2 y de las pseudosemimétricas ϕ_1 , ϕ_3 , ϕ_4 y ψ_2 . Estas constantes y funciones deberán ser determinadas por el/la investigador/a al comiendo del estudio, dependiendo de sus intereses, o en función de la clase de estudio que se quiere desarrollar. La principal ventaja de D es la enorme clase de particularizaciones que se pueden obtener sobre el conjunto \mathcal{F} en toda su extensión, y no sólo sobre los números difusos triangulares o trapezoidales.

Las funciones h_1 y h_2 reciben el siguiente nombre en el ambiente difuso.

Definición 3.3.5 Una negación difusa es una función no creciente $h : \mathbb{I} \rightarrow \mathbb{I}$ tal que $h(0) = 1$ y $h(1) = 0$.

Ejemplos de negaciones difusas continuas son la *negación estándar* $h(\alpha) = 1 - \alpha$, la *negación coseno* $h(\alpha) = (1 + \cos(\pi\alpha))/2$, la *negación de Sugeno* $h(\alpha) = (1 - \alpha)/(1 + \lambda\alpha)$ y la *negación de Yager* $h(\alpha) = (1 - \alpha^\lambda)^{1/\lambda}$.

Antes de continuar, presentamos, en la Tabla 3.1, algunos ejemplos que muestran cómo actúa la función D en diferentes casos: en las primeras columnas, la distancia es necesariamente un número triangular o trapezoidal (cuando $h_1(\alpha) = h_2(\alpha) = 1 - \alpha$ para todo $\alpha \in \mathbb{I}$); sin embargo, en la última columna, el investigador desea considerar formas más complejas.

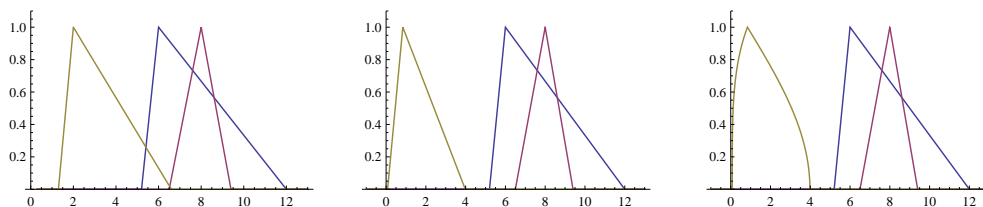


Figura 3.1: Acción de la aplicación D utilizando diferentes funciones h_1 y h_2 (Tabla 3.1), siendo $\mathcal{A} = (5.2/6/12)$ y $\mathcal{B} = (6.5/8/9.4)$.

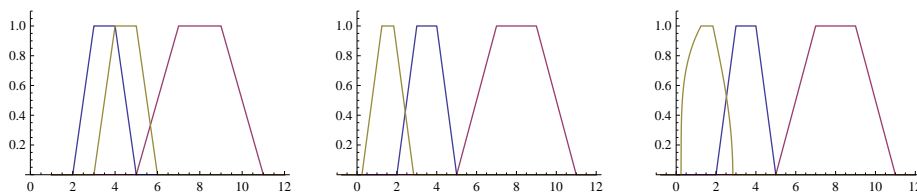


Figura 3.2: Acción de la aplicación D utilizando diferentes funciones h_1 y h_2 (Tabla 3.1), siendo $\mathcal{A} = (2/3/4/5)$ y $\mathcal{B} = (5/7/9/11)$.

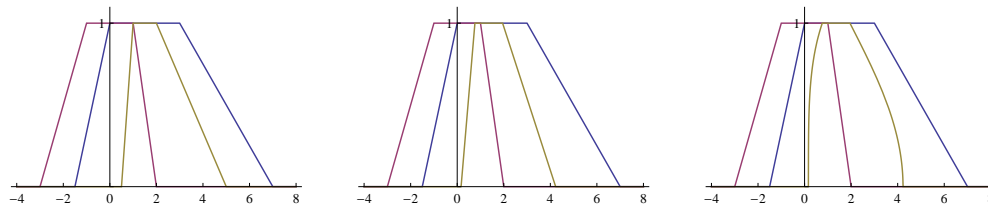


Figura 3.3: Acción de la aplicación D utilizando diferentes funciones h_1 y h_2 (Tabla 3.1), siendo $\mathcal{A} = (-1.5/0/3/7)$ y $\mathcal{B} = (-3/ -1/1/2)$.

Tabla 3.1: Algunas medidas $D(\mathcal{A}, \mathcal{B})$, donde $\mathcal{D} = \mathcal{D}_c$, $d = \delta$, $q_1 = q_2 = 1$ y $\phi_i(x, y) = \psi_2(x, y) = |x - y|^p$.

Número difuso	$p = 1$	$p = 0.75$
trapezoidal	$h_1(\alpha) = h_2(\alpha) = 1 - \alpha$	
$A = (5.2/6/12)$ $B = (6.5/8/9.4)$	(1.3/2/6.6)	(0.076/0.84/3.98)
$A = (2/3/4/5)$ $B = (5/7/9/11)$	(3/4/5/6)	(0.248/1.248/1.842/2.842)
$A = (-1.5/0/3/7)$ $B = (-3/-1/1/2)$	(0.5/1/2/5)	(0.166/0.761/1.95/4.23)
Número difuso	$p = 0.75$	
trapezoidal	$h_1(\alpha) = 1 - \alpha^4$, $h_2(\alpha) = \cos(\pi\alpha/2)$	
$A = (5.2/6/12)$ $B = (6.5/8/9.4)$	$L(t) \approx 0.9292\sqrt[4]{1.752t - 0.1325}$, $0.076 \leq t \leq 0.84$ $R(t) \approx (2/\pi) \arccos(0.318t - 0.268)$, $0.84 \leq t \leq 3.98$	
$A = (2/3/4/5)$ $B = (5/7/9/11)$	$L(t) \approx 0.707\sqrt[4]{4x - 0.99}$, $0.248 \leq t \leq 1.248$ $R(t) \approx (2/\pi) \arccos(x - 1.842)$, $1.842 \leq t \leq 2.842$	
$A = (-1.5/0/3/7)$ $B = (-3/-1/1/2)$	$L(t) \approx 0.993\sqrt[4]{1.733x - 0.288}$, $0.166 \leq t \leq 0.761$ $R(t) \approx (2/\pi) \arccos(0.439t - 0.855)$, $1.95 \leq t \leq 4.23$	

La Figura 3.1 muestra el efecto de las funciones h_1 y h_2 sobre la medida de distancia entre dos números trapezoidales \mathcal{A} (representado en color azul) y \mathcal{B} (representado en color violeta). El número $D(\mathcal{A}, \mathcal{B})$ se ha dibujado en color verde. Las dos primeras figuras muestran que, como demostraremos más ade-

lante, el resultado es, de nuevo, un número difuso trapezoidal cuando tomamos $h_1(\alpha) = h_2(\alpha) = 1 - \alpha$. Cuando las funciones h_1 y h_2 no son las mencionadas anteriormente, el resultado no es necesariamente un número difuso trapezoidal. Éste es el caso del tercer gráfico de la Figura 3.1, en el que la medida de distancia entre dos números difusos trapezoidales es un número difuso de tipo LR. Las Figuras 3.2 y 3.3 representan este efecto de la medida de similitud D sobre dos números difusos trapezoidales diferentes y se interpretan de la misma forma que la Figura 3.1.

Observación 3.3.6 *Nótese que en la Definición 3.3.4 hemos utilizado la métrica $\delta : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_{0_{\mathcal{I}}, \subseteq}^+$ definida por la expresión (3.4). En realidad, se podría haber utilizado cualquier otra métrica $d : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_{0_{\mathcal{I}}, \subseteq}^+$. No obstante, de cara a las aplicaciones prácticas que consideraremos, nos basta con utilizar δ . Se deja para estudios posteriores la posibilidad de utilizar otra métrica d sobre $(\mathcal{I}, 0_{\mathcal{I}}, \subseteq, s_{\mathcal{I}})$.*

Observación 3.3.7 *Observemos que la forma geométrica de $D(\mathcal{A}, \mathcal{B})$ está determinada no sólo por h_1 y h_2 , sino también por $\delta(\ker \mathcal{A}, \text{supp } \mathcal{A})$, por $\delta(\ker \mathcal{B}, \text{supp } \mathcal{B})$ y por las pseudosemimétricas ϕ_3 y ϕ_4 .*

Observación 3.3.8 *Teniendo en cuenta las aplicaciones, estamos muy interesados en la defusificación \mathcal{D}_c , las funciones $h_i(\alpha) = 1 - \alpha$ para cualquier $\alpha \in \mathbb{I}$, la distancia intervalar δ con respecto a la inclusión \subseteq y el orden parcial \preceq sobre \mathcal{T} . Sin embargo, muchas propiedades son ciertas en un contexto más general. Por lo tanto, desarrollaremos los aspectos teóricos en el caso más general posible, utilizando medidas de similitud arbitrarias.*

Ejemplo 3.3.9 *Si $q_2 = 0$ y $h_1(\alpha) = h_2(\alpha) = 0$ para todo $\alpha \in \mathbb{I}$, entonces, para cada $\mathcal{A}, \mathcal{B} \in \mathcal{F}$, $D(\mathcal{A}, \mathcal{B})$ es el número difuso crisp $q_1 \phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B})$. Por ejemplo, podemos obtener*

$$q_1 |(\mathcal{D}\mathcal{A})^{k_1} - (\mathcal{D}\mathcal{B})^{k_1}|^{p_1}$$

donde $q_1, k_1 > 0$. En tal caso, si consideramos los números difusos crisp $\mathcal{A} = \tilde{a}$

y $\mathcal{B} = \tilde{b}$, $q_1 = k_1 = p_1 = 1$ y $\mathcal{D}\tilde{a} = a$, entonces D es, sobre el conjunto de los números difusos crisp, la métrica euclídea $D(\tilde{a}, \tilde{b}) = |a - b|$.

A lo largo del resto de esta sección, vamos a probar que las consideraciones geométricas que nos han llevado a introducir la Definición 3.3.4 se cumplen bajo hipótesis adicionales sobre las variables $q_1, q_2, \mathcal{D}, h_1, h_2, \phi_1, \phi_3, \phi_4$ y ψ_2 .

Lema 3.3.10 Para todo $\mathcal{A}, \mathcal{B} \in \mathcal{F}$,

$$D(\mathcal{B}, \mathcal{A}) = D(\mathcal{A}, \mathcal{B}) \quad \text{y} \quad D(\mathcal{A}, \mathcal{A}) = \tilde{0},$$

es decir, D es una pseudosemimétrica sobre $(\mathcal{F}, \tilde{0})$.

DEMOSTRACIÓN: Como ϕ_1, ϕ_3, ϕ_4 y ψ_2 son pseudosemimétricas, entonces

$$\phi(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{A}) = 0 \quad \text{y} \quad \phi(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) = \phi(\mathcal{D}\mathcal{B}, \mathcal{D}\mathcal{A})$$

para cada $\phi \in \{\phi_1, \phi_3, \phi_4, \psi_2\}$ y cada $\mathcal{A}, \mathcal{B} \in \mathcal{F}$. Como consecuencia, para cada $\alpha \in \mathbb{I}$,

$$\begin{aligned} \underline{D(\mathcal{A}, \mathcal{A})}_\alpha &= q_1 \phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{A}) - q_2 \psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{A}) \\ &\quad - h_1(\alpha) \phi_3(\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A})) \\ &= q_1 0 - q_2 0 - h_1(\alpha) 0 = 0, \end{aligned}$$

y, de igual forma, $\overline{D(\mathcal{A}, \mathcal{A})}_\alpha = 0$. Por lo tanto, $D(\mathcal{A}, \mathcal{A}) = \tilde{0}$. Además, como δ es también simétrica,

$$\begin{aligned} \underline{D(\mathcal{B}, \mathcal{A})}_\alpha &= q_1 \phi_1(\mathcal{D}\mathcal{B}, \mathcal{D}\mathcal{A}) - q_2 \psi_2(\text{spr } \mathcal{B}, \text{spr } \mathcal{A}) \\ &\quad - h_1(\alpha) \phi_3(\underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B}), \underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A})) \\ &= q_1 \phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) - q_2 \psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) \\ &\quad - h_1(\alpha) \phi_3(\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})) \\ &= \underline{D(\mathcal{A}, \mathcal{B})}_\alpha, \end{aligned}$$

y, de la misma forma, $\overline{D(\mathcal{B}, \mathcal{A})}_\alpha = \overline{D(\mathcal{A}, \mathcal{B})}_\alpha$ para cada $\alpha \in \mathbb{I}$. Dicho de otro modo, $D(\mathcal{B}, \mathcal{A}) = D(\mathcal{A}, \mathcal{B})$. ■

El siguiente teorema muestra que, bajo ciertas condiciones, algunos de los subconjuntos más útiles de \mathcal{F} (incluido el conjunto \mathcal{T} de los números trapezoidales) son cerrados bajo D . Esta propiedad garantiza que se puede elegir, a priori, el rango de la métrica según los intereses y la profundidad del estudio: por ejemplo, si consideramos que los números difusos triangulares son adecuados para explicar los fenómenos del experimento que se considere, tiene sentido imponer desde el principio que la medida de similitud entre ambos también sea un número difuso triangular.

Teorema 3.3.11 Si se toman $h_1(\alpha) = h_2(\alpha) = 1 - \alpha$ para cada $\alpha \in \mathbb{I}$ en (3.8)-(3.9) y \mathcal{A} y \mathcal{B} son números difusos trapezoidales (respectivamente, triangulares, rectangulares, reales), entonces $D(\mathcal{A}, \mathcal{B})$ es también un número difuso trapezoidal (respectivamente, triangular, rectangular, real).

De esta forma, si $q_1, q_2 > 0$, $\mathcal{D} = \mathcal{D}_c$ y ϕ_1, ψ_2, ϕ_3 y ϕ_4 son semimétricas, entonces la aplicación $D : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$ es una semimétrica sobre $(\mathcal{T}, \tilde{0}, \preceq)$.

DEMOSTRACIÓN: Supongamos que $\mathcal{A} = \text{Tra}(A^c, A^m, A^\ell, A^s)$ y $\mathcal{B} = \text{Tra}(B^c, B^m, B^\ell, B^s)$ son dos números trapezoidales arbitrarios. Entonces

$$\begin{aligned} \delta(\ker \mathcal{A}, \text{supp } \mathcal{A}) &= \delta([A^c - A^m, A^c + A^m], [A^c - A^m - A^\ell, A^c + A^m + A^s]) \\ &= [-|(A^c - A^m) - (A^c - A^m - A^\ell)|, \\ &\quad |(A^c + A^m) - (A^c + A^m + A^s)|] \\ &= [-A^\ell, A^s]. \end{aligned}$$

Por consiguiente, para cada $\mathcal{A} \in \mathcal{T}$,

$$\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}) = -A^\ell \quad \text{y} \quad \overline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}) = A^s.$$

Dado que $\text{spr } \mathcal{A} = A^m$ y $\text{spr } \mathcal{B} = B^m$, entonces las expresiones (3.8)-(3.9) quedan de la siguiente manera:

$$\begin{aligned} \underline{D(\mathcal{A}, \mathcal{B})}_\alpha &= q_1 \phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) - q_2 \psi_2(A^m, B^m) - (1 - \alpha) \phi_3(-A^\ell, -B^\ell), \\ \overline{D(\mathcal{A}, \mathcal{B})}_\alpha &= q_1 \phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) + q_2 \psi_2(A^m, B^m) + (1 - \alpha) \phi_4(A^s, B^s). \end{aligned}$$

De esta forma, $D(\mathcal{A}, \mathcal{B})$ es el número trapezoidal

$$D(\mathcal{A}, \mathcal{B}) = \text{Tra} \left(q_1 \phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}), q_2 \psi_2(A^m, B^m), \phi_3(-A^\ell, -B^\ell), \phi_4(A^s, B^s) \right) \quad (3.10)$$

para cada $\mathcal{A}, \mathcal{B} \in \mathcal{T}$, lo que demuestra que $D(\mathcal{A}, \mathcal{B})$ también es un número difuso trapezoidal. Así, la aplicación $D : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$ está bien definida. Además, se tiene lo siguiente.

- Si \mathcal{A} y \mathcal{B} son números difusos triangulares, entonces $A^m = B^m = 0$, de donde $D(\mathcal{A}, \mathcal{B})^m = q_2 \psi_2(A^m, B^m) = 0$. Así, $D(\mathcal{A}, \mathcal{B})$ también es triangular.
- Si \mathcal{A} y \mathcal{B} son números difusos rectangulares, entonces $A^\ell = A^s = B^\ell = B^s = 0$, de donde

$$\begin{aligned} D(\mathcal{A}, \mathcal{B})^\ell &= \phi_3(-A^\ell, -B^\ell) = \phi_3(0, 0) = 0 \quad \text{y} \\ D(\mathcal{A}, \mathcal{B})^s &= \phi_4(A^s, B^s) = \phi_4(0, 0) = 0. \end{aligned}$$

Por tanto, $D(\mathcal{A}, \mathcal{B})$ también es rectangular.

- Si \mathcal{A} y \mathcal{B} son números difusos reales, entonces $A^m = A^\ell = A^s = B^m = B^\ell = B^s = 0$, de donde $D(\mathcal{A}, \mathcal{B})^m = D(\mathcal{A}, \mathcal{B})^\ell = D(\mathcal{A}, \mathcal{B})^s = 0$. Así, $D(\mathcal{A}, \mathcal{B})$ también es real.

Supongamos ahora que $q_1, q_2 > 0$ y que $\mathcal{D} = \mathcal{D}_c$. El Lema 3.3.10 nos asegura que D es una pseudosemimétrica sobre $(\mathcal{F}, \tilde{0})$. Además, si ocurriese que $D(\mathcal{A}, \mathcal{B}) = 0$ para ciertos $\mathcal{A}, \mathcal{B} \in \mathcal{T}$, la expresión (3.10) nos garantizaría que

$$q_1 \phi_1(\mathcal{D}_c \mathcal{A}, \mathcal{D}_c \mathcal{B}) = q_2 \psi_2(A^m, B^m) = \phi_3(-A^\ell, -B^\ell) = \phi_4(A^s, B^s) = 0.$$

Como $q_1, q_2 > 0$ y ϕ_1, ψ_2, ϕ_3 y ϕ_4 son semimétricas, deducimos que $A^c = \mathcal{D}_c \mathcal{A} = \mathcal{D}_c \mathcal{B} = B^c$, $A^m = B^m$, $A^\ell = B^\ell$ y $A^s = B^s$. De esta forma, $\mathcal{A} = \mathcal{B}$, lo que concluye que D es una semimétrica sobre $(\mathcal{F}, \tilde{0})$. ■

Observación 3.3.12 *Una de las elecciones más naturales que puede hacer un/a investigador/a al desarrollar un proceso de regresión difusa consiste en tomar*

$$\begin{aligned} h_1(\alpha) &= h_2(\alpha) = 1 - \alpha \quad \text{para cada } \alpha \in \mathbb{I}, \\ \left. \begin{aligned} \phi_1(x, y) &= q_1 |x^{k_1} - y^{k_1}|^{p_1} \\ \phi_3(x, y) &= q_3 |x^{k_3} - y^{k_3}|^{p_3}, \\ \phi_4(x, y) &= q_4 |x^{k_4} - y^{k_4}|^{p_4} \end{aligned} \right\} \quad \text{para cada } x, y \in \mathbb{R}, \\ \psi_2(x, y) &= q_2 |x^{k_2} - y^{k_2}|^{p_2} \quad \text{para cada } x, y \in \mathbb{R}_0^+, \end{aligned}$$

en las expresiones (3.8)-(3.9), donde $q_i, k_i, p_i > 0$ para cada $i \in \{1, 2, 3, 4\}$. En tal caso, el Lema 3.1.6 nos asegura que ϕ_1 , ψ_2 , ϕ_3 y ϕ_4 son semimétricas. En tal caso, la expresión (3.10) nos asegura que

$$\begin{aligned} D(\mathcal{A}, \mathcal{B}) = \text{Tra} \left(q_1 \left| (\mathcal{D}\mathcal{A})^{k_1} - (\mathcal{D}\mathcal{B})^{k_1} \right|^{p_1}, q_2 \left| (A^m)^{k_2} - (B^m)^{k_2} \right|^{p_2}, \right. \\ \left. q_3 \left| (A^\ell)^{k_3} - (B^\ell)^{k_3} \right|^{p_3}, q_4 \left| (A^s)^{k_4} - (B^s)^{k_4} \right|^{p_4} \right) \quad (3.11) \end{aligned}$$

para cada $\mathcal{A} = \text{Tra}(A^c, A^m, A^\ell, A^s)$ y $\mathcal{B} = \text{Tra}(B^c, B^m, B^\ell, B^s)$.

Además, en el caso de que se elija $\mathcal{D} = \mathcal{D}_c$, tendremos que

$$\begin{aligned} D(\mathcal{A}, \mathcal{B}) = \text{Tra} \left(q_1 \left| (A^c)^{k_1} - (B^c)^{k_1} \right|^{p_1}, q_2 \left| (A^m)^{k_2} - (B^m)^{k_2} \right|^{p_2}, \right. \\ \left. q_3 \left| (A^\ell)^{k_3} - (B^\ell)^{k_3} \right|^{p_3}, q_4 \left| (A^s)^{k_4} - (B^s)^{k_4} \right|^{p_4} \right), \quad (3.12) \end{aligned}$$

lo que nos dice que, bajo ciertas condiciones, la medida de similitud D opera componente a componente cuando los números difusos trapezoidales se expresan de la forma $\mathcal{A} = \text{Tra}(A^c, A^m, A^\ell, A^s)$.

En este caso, el orden parcial \preceq también se tomaría componente a componente ya que

$$\begin{aligned} \mathcal{A} \preceq \mathcal{B} \\ \Leftrightarrow [\mathcal{D}_c \mathcal{A} \leq \mathcal{D}_c \mathcal{B}, \text{spr } \mathcal{A} \leq \text{spr } \mathcal{B}, \delta(\ker \mathcal{A}, \text{supp } \mathcal{A}) \subseteq \delta(\ker \mathcal{B}, \text{supp } \mathcal{B})] \\ \Leftrightarrow \{ A^c \leq B^c, A^m \leq B^m, [-A^\ell, A^s] \subseteq [-B^\ell, B^s] \} \\ \Leftrightarrow \{ A^c \leq B^c, A^m \leq B^m, A^\ell \leq B^\ell, A^s \leq B^s \}. \quad (3.13) \end{aligned}$$

En [52], Roldán *et al.* hicieron un completo estudio de los números difusos cuya imagen es un subconjunto finito o infinito numerable de \mathbb{I} . También en este caso se puede conseguir que la medida de similitud tome estos valores difusos.

Teorema 3.3.13 Si h_1 y h_2 tienen imagen finita (respectivamente, numerable), entonces $D(\mathcal{A}, \mathcal{B})$ es un número difuso de imagen finita (respectivamente, numerable).

Los siguientes resultados muestran algunas propiedades métricas de la distancia D que nos permiten afirmar que, en algunos casos, D es una métrica sobre $(\mathcal{T}, \tilde{0}, +)$ con respecto al orden parcial \preceq .

Lema 3.3.14 Si $h_1(1) = h_2(1) = 0$, entonces la medida de distancia D verifica las siguientes propiedades.

1. $\ker D(\mathcal{A}, \mathcal{B}) = [q_1\phi_1(\mathcal{DA}, \mathcal{DB}) - q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}), q_1\phi_1(\mathcal{DA}, \mathcal{DB}) + q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B})]$.
2. $\mathcal{D}_c(D(\mathcal{A}, \mathcal{B})) = q_1\phi_1(\mathcal{DA}, \mathcal{DB})$.
3. $\text{spr } D(\mathcal{A}, \mathcal{B}) = q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B})$.
4. $D(\mathcal{A}, \mathcal{B}) \succcurlyeq \tilde{0}$ con respecto a $(\mathcal{D}_c, \delta, \subseteq)$.

DEMOSTRACIÓN: Como $h_1(1) = h_2(1) = 0$, es claro que:

$$\begin{aligned} \ker D(\mathcal{A}, \mathcal{B}) &= [\underline{D(\mathcal{A}, \mathcal{B})}_1, \overline{D(\mathcal{A}, \mathcal{B})}_1] \\ &= [q_1\phi_1(\mathcal{DA}, \mathcal{DB}) - q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}), \\ &\quad q_1\phi_1(\mathcal{DA}, \mathcal{DB}) + q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B})]. \end{aligned}$$

(2) A partir de (1):

$$\begin{aligned}\mathcal{D}_c(D(\mathcal{A}, \mathcal{B})) &= \frac{D(\mathcal{A}, \mathcal{B})_1 + \overline{D(\mathcal{A}, \mathcal{B})}_1}{2} \\ &= \frac{1}{2} \left(q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) - q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) + \right. \\ &\quad \left. q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) + q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) \right) \\ &= q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}).\end{aligned}$$

(3) Similarmente:

$$\begin{aligned}\text{spr } D(\mathcal{A}, \mathcal{B}) &= \frac{\overline{D(\mathcal{A}, \mathcal{B})}_1 - D(\mathcal{A}, \mathcal{B})_1}{2} \\ &= \frac{1}{2} \left(q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) + q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) - \right. \\ &\quad \left. (q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) - q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B})) \right) \\ &= q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}).\end{aligned}$$

(4) Si $\mathcal{D} = \mathcal{D}_c$, entonces el apartado 2 muestra que

$$\mathcal{D}_c(D(\mathcal{A}, \mathcal{B})) = q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) \geq 0.$$

Así, el apartado 2 del Lema 3.3.2 garantiza que $D(\mathcal{A}, \mathcal{B}) \succcurlyeq \tilde{0}$ con respecto a $(\mathcal{D}_c, \delta, \subseteq)$. ■

Teorema 3.3.15 Si $h_1(1) = h_2(1) = 0$ y ϕ_1, ϕ_3, ϕ_4 y ψ_2 son pseudométricas, entonces

$$D(\mathcal{A}, \mathcal{C}) \preccurlyeq D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{A})$$

para cada $\mathcal{A}, \mathcal{B}, \mathcal{C} \in \mathcal{F}$.

DEMOSTRACIÓN: Consideremos la terna $(\mathcal{D}_c, \delta, \subseteq)$ y vamos a probar las tres condiciones que verifican las pseudométricas. Por un lado, por el Lema 2.2.13 y el apartado 2 del Lema 3.3.14,

$$\begin{aligned}\mathcal{D}_c(D(\mathcal{A}, \mathcal{C})) &= q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{C}) \leq q_1\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) + q_1\phi_1(\mathcal{D}\mathcal{B}, \mathcal{D}\mathcal{C}) \\ &= \mathcal{D}_c(D(\mathcal{A}, \mathcal{B})) + \mathcal{D}_c(D(\mathcal{B}, \mathcal{C})) = \mathcal{D}_c(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})).\end{aligned}$$

Por otro lado, utilizando el Lema 3.2.3 y el apartado 3 del Lema 3.3.14,

$$\begin{aligned} \text{spr } D(\mathcal{A}, \mathcal{C}) &= q_2 \psi_2 (\text{spr } \mathcal{A}, \text{spr } \mathcal{C}) \leq q_2 \psi_2 (\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) + q_2 \psi_2 (\text{spr } \mathcal{B}, \text{spr } \mathcal{C}) \\ &= \text{spr } D(\mathcal{A}, \mathcal{B}) + \text{spr } D(\mathcal{B}, \mathcal{C}) = \text{spr}(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})). \end{aligned}$$

Finalmente, vamos a probar que:

$$\begin{aligned} \delta(\ker D(\mathcal{A}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{C})) &\subseteq \\ &\delta(\ker(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})), \text{supp}(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C}))). \end{aligned}$$

Según el Lema 3.2.3, el apartado 2 del Lema 3.2.2, y teniendo en cuenta que ϕ_3 y ϕ_4 verifican la desigualdad triangular, deducimos que:

$$\begin{aligned} \underline{\delta}(\ker D(\mathcal{A}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{C})) &= -h_1(0) \phi_3 (\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{\delta}(\ker \mathcal{C}, \text{supp } \mathcal{C})) \\ &\geq -h_1(0) \phi_3 (\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})) \\ &\quad - h_1(0) \phi_3 (\underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B}), \underline{\delta}(\ker \mathcal{C}, \text{supp } \mathcal{C})) \\ &= \underline{\delta}(\ker D(\mathcal{A}, \mathcal{B}), \text{supp } D(\mathcal{A}, \mathcal{B})) + \underline{\delta}(\ker D(\mathcal{B}, \mathcal{C}), \text{supp } D(\mathcal{B}, \mathcal{C})) \\ &= \underline{\delta}(\ker D(\mathcal{A}, \mathcal{B}) + \ker D(\mathcal{B}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{B}) + \text{supp } D(\mathcal{B}, \mathcal{C})) \\ &= \underline{\delta}(\ker(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})), \text{supp}(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C}))) \end{aligned}$$

y, de la misma forma,

$$\begin{aligned} \overline{\delta}(\ker D(\mathcal{A}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{C})) &= h_2(0) \phi_4 (\overline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \overline{\delta}(\ker \mathcal{C}, \text{supp } \mathcal{C})) \\ &\leq h_2(0) \phi_4 (\overline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \overline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})) \\ &\quad + h_2(0) \phi_4 (\overline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B}), \overline{\delta}(\ker \mathcal{C}, \text{supp } \mathcal{C})) \\ &= \overline{\delta}(\ker D(\mathcal{A}, \mathcal{B}), \text{supp } D(\mathcal{A}, \mathcal{B})) + \overline{\delta}(\ker D(\mathcal{B}, \mathcal{C}), \text{supp } D(\mathcal{B}, \mathcal{C})) \\ &= \overline{\delta}(\ker D(\mathcal{A}, \mathcal{B}) + \ker D(\mathcal{B}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{B}) + \text{supp } D(\mathcal{B}, \mathcal{C})) \\ &= \overline{\delta}(\ker(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})), \text{supp}(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C}))). \end{aligned}$$

Por lo tanto:

$$\begin{aligned}
& \delta(\ker D(\mathcal{A}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{C})) \\
&= [\underline{\delta}(\ker D(\mathcal{A}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{C})), \bar{\delta}(\ker D(\mathcal{A}, \mathcal{C}), \text{supp } D(\mathcal{A}, \mathcal{C}))] \\
&\subseteq [\underline{\delta}(\ker(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})), \text{supp}(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C}))), \\
&\quad \bar{\delta}(\ker(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})), \text{supp}(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})))] \\
&= \delta(\ker(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C})), \text{supp}(D(\mathcal{A}, \mathcal{B}) + D(\mathcal{B}, \mathcal{C}))).
\end{aligned}$$

■

El resultado más importante de esta sección nos permite asegurar que, en algunos casos, la medida de similitud $D : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{F}$ es una métrica en el conjunto de los números difusos en el sentido de la Definición 3.1.2.

Teorema 3.3.16 Si $q_1, q_2 > 0$, $h_1(\alpha) = h_2(\alpha) = 1 - \alpha$ para cada $\alpha \in \mathbb{I}$, $\mathcal{D} = \mathcal{D}_c$ y consideramos la distancia intervalar δ sobre \mathcal{I} con respecto al orden parcial \subseteq , entonces D es una métrica sobre $(\mathcal{T}, \tilde{0}, +)$ con respecto al orden parcial \preceq , cualesquiera que sean las métricas ϕ_1, ϕ_3, ϕ_4 y ψ_2 .

DEMOSTRACIÓN: El Teorema 3.3.11 nos permite afirmar que si $\mathcal{A}, \mathcal{B} \in \mathcal{T}$, entonces $D(\mathcal{A}, \mathcal{B}) \in \mathcal{T}$. El Lema 3.3.2 muestra que \preceq es un orden parcial sobre \mathcal{T} . El apartado 4 del Lema 3.3.14 nos permite afirmar que $D(\mathcal{A}, \mathcal{B}) \succ \tilde{0}$. El Lema 3.3.10 nos garantiza que D es una pseudosemimétrica sobre $(\mathcal{T}, \tilde{0})$. El Teorema 3.3.15 nos dice que D verifica la desigualdad triangular. Finalmente, si $D(\mathcal{A}, \mathcal{B}) = \tilde{0}$, entonces

$$\begin{aligned}
\phi_1(\mathcal{D}\mathcal{A}, \mathcal{D}\mathcal{B}) &= \psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) \\
&= \phi_3(\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})) \\
&= \phi_4(\bar{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \bar{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})) = 0.
\end{aligned}$$

Como ϕ_1, ψ_2, ϕ_3 y ϕ_4 son métricas, deducimos que $\mathcal{D}\mathcal{A} = \mathcal{D}\mathcal{B}$, $\text{spr } \mathcal{A} = \text{spr } \mathcal{B}$ y $\underline{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}) = \underline{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})$, es decir, $\mathcal{A} \preceq \mathcal{B}$ y $\mathcal{B} \preceq \mathcal{A}$ y, por lo tanto, $\mathcal{A} = \mathcal{B}$.

■

3.4. Aplicaciones

Disponer de una métrica sobre un conjunto permite desarrollar una serie de propiedades que no serían posibles sin la distancia. Aunque nuestra métrica toma valores difusos, vamos a deducir algunas consecuencias directas de la forma en la que medimos la similitud entre dos números difusos.

3.4.1. La topología inducida sobre \mathcal{T}

Una métrica real sobre un conjunto arbitrario induce, de manera natural, una topología Hausdorff. Veamos, a continuación, que la métrica difusa D también induce, de manera natural, una topología sobre \mathcal{T} . Para ello, demostramos la siguiente propiedad.

Proposición 3.4.1 Bajo la hipótesis del Teorema 3.3.16, si $\mathcal{A}, \mathcal{B} \in \mathcal{T}$, y $D(\mathcal{A}, \mathcal{B})$ es un número difuso crisp, entonces existe $r \in \mathbb{R}$ tal que $\mathcal{A} = r + \mathcal{B}$, es decir, \mathcal{A} es una traslación de \mathcal{B} .

DEMOSTRACIÓN: Supongamos que $\mathcal{A} = (a/b/c/e)$ y $\mathcal{B} = (a'/b'/c'/e')$. Si $D(\mathcal{A}, \mathcal{B})$ es un número difuso crisp, entonces

$$\begin{aligned} \psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) &= \phi_3(\underline{d}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \underline{d}(\ker \mathcal{B}, \text{supp } \mathcal{B})) \\ &= \phi_4(\bar{\delta}(\ker \mathcal{A}, \text{supp } \mathcal{A}), \bar{\delta}(\ker \mathcal{B}, \text{supp } \mathcal{B})) = 0. \end{aligned}$$

Esto significa que $\text{spr } \mathcal{A} = \text{spr } \mathcal{B}$ y $\delta(\ker \mathcal{A}, \text{supp } \mathcal{A}) = \delta(\ker \mathcal{B}, \text{supp } \mathcal{B})$, es decir,

$$\frac{c-b}{2} = \frac{c'-b'}{2}, \quad b-a = b'-a' \quad \text{y} \quad e-c = e'-c'.$$

Si llamamos $r = a - a'$, resulta que $b - b' = c - c' = e - e' = r$, por lo que $a = r + a'$, $b = r + b'$, $c = r + c'$ y $e = r + e'$. ■

Esto nos permite considerar una topología sobre \mathcal{T} como se muestra a continuación.

Teorema 3.4.2 Bajo la hipótesis del Teorema 3.3.16, consideremos, para cada $\mathcal{A} \in \mathcal{T}$ y cada $r > 0$, el conjunto

$$N_r(\mathcal{A}) = \{\mathcal{B} \in \mathcal{T} : D(\mathcal{A}, \mathcal{B}) \prec \tilde{r}\}.$$

Entonces, la familia $\{N_r(\mathcal{A})\}_{r>0}$ es un sistema de entornos de \mathcal{A} en una topología Hausdorff sobre \mathcal{T} que verifica el primer axioma de numerabilidad.

DEMOSTRACIÓN: En primer lugar, vamos a caracterizar:

$$N_r(\mathcal{A}) = \{\mathcal{B} \in \mathcal{T} : \mathcal{B} = (\mathcal{D}_c\mathcal{B} - \mathcal{D}_c\mathcal{A}) + \mathcal{A} \text{ y } q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) < r \}.$$

En efecto, sea $\mathcal{B} \in N_r(\mathcal{A})$. Como $D(\mathcal{A}, \mathcal{B}) \prec \tilde{r}$, Entonces

- $q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) = \mathcal{D}_c(D(\mathcal{A}, \mathcal{B})) \leq \mathcal{D}_c\tilde{r} = r$;
- $0 \leq q_2\psi_2(\text{spr } \mathcal{A}, \text{spr } \mathcal{B}) = \text{spr } D(\mathcal{A}, \mathcal{B}) \leq \text{spr } \tilde{r} = 0 \Rightarrow \text{spr } \mathcal{A} = \text{spr } \mathcal{B}$;
- $0_{\mathcal{I}} \subseteq \delta(\ker D(\mathcal{A}, \mathcal{B}), \text{supp } D(\mathcal{A}, \mathcal{B})) \subseteq \delta(\ker \tilde{r}, \text{supp } \tilde{r}) = \delta([r, r], [r, r]) = 0_{\mathcal{I}}$
 $\Rightarrow \ker D(\mathcal{A}, \mathcal{B}) = \text{supp } D(\mathcal{A}, \mathcal{B})$
 $\Rightarrow D(\mathcal{A}, \mathcal{B})$ es rectangular.

Como $D(\mathcal{A}, \mathcal{B})$ es rectangular, pero $\text{spr } D(\mathcal{A}, \mathcal{B}) = 0$, entonces $D(\mathcal{A}, \mathcal{B})$ es un número difuso crisp. En este caso, la proposición anterior demuestra que \mathcal{B} es una translación de \mathcal{A} , es decir, existe $s \in \mathbb{R}$ tal que $\mathcal{B} = s + \mathcal{A}$. Considerando la defusificación \mathcal{D}_c , observamos que $\mathcal{D}_c\mathcal{B} = \mathcal{D}_c(s + \mathcal{A}) = \mathcal{D}_c s + \mathcal{D}_c\mathcal{A} = s + \mathcal{D}_c\mathcal{A}$, y, por lo tanto, $s = \mathcal{D}_c\mathcal{B} - \mathcal{D}_c\mathcal{A}$. Si $q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) = r$, entonces $D(\mathcal{A}, \mathcal{B}) \preceq \tilde{r}$ y $\tilde{r} \preceq D(\mathcal{A}, \mathcal{B})$, por lo que $D(\mathcal{A}, \mathcal{B}) = \tilde{r}$, pero suponemos que $D(\mathcal{A}, \mathcal{B}) \prec \tilde{r}$. Entonces, $q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) < r$.

A continuación, veamos que $\beta_{\mathcal{A}} = \{N_r(\mathcal{A})\}_{r>0}$ es un sistema de entornos de

\mathcal{A} . Tenemos que probar tres propiedades.

- $\mathcal{A} \in U$ para cada $U \in \beta_{\mathcal{A}}$.
- Para cada $U, V \in \beta_{\mathcal{A}}$, existe $W \in \beta_{\mathcal{A}}$ tal que $W \subseteq U \cap V$.
- Para cada $U \in \beta_{\mathcal{A}}$, existe $U_0 \in \beta_{\mathcal{A}}$ tal que para cada $\mathcal{B} \in U_0$ existe $V \in \beta_{\mathcal{B}}$ tal que $V \subseteq U$.

Las dos primeras son triviales, ya que $\mathcal{A} \in N_r(\mathcal{A})$ para todo $r > 0$ y $N_{r_3}(\mathcal{A}) \subseteq N_{r_1}(\mathcal{A}) \cap N_{r_2}(\mathcal{A})$ si $0 < r_3 < \min(r_1, r_2)$. Sea $U = U_0 = N_r(\mathcal{A})$ y sea $\mathcal{B} \in N_r(\mathcal{A})$. Si $\mathcal{B} = \mathcal{A}$, podemos tomar $V = U$. Supongamos que $\mathcal{B} \neq \mathcal{A}$. Como $q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) < r$, sean $s, \varepsilon > 0$ tal que $q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) < s < s + \varepsilon < r$. Consideremos $V = N_\varepsilon(\mathcal{B}) \in \beta_{\mathcal{B}}$ y veamos que $N_\varepsilon(\mathcal{B}) \subseteq N_r(\mathcal{A})$. En efecto, sea $\mathcal{C} \in N_\varepsilon(\mathcal{B})$. Sabemos que $\mathcal{C} = (\mathcal{D}_c\mathcal{C} - \mathcal{D}_c\mathcal{B}) + \mathcal{B}$ y $q_1\phi_1(\mathcal{D}_c\mathcal{B}, \mathcal{D}_c\mathcal{C}) < \varepsilon$. Entonces:

$$\begin{aligned} q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{C}) &\leq q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) + q_1\phi_1(\mathcal{D}_c\mathcal{B}, \mathcal{D}_c\mathcal{C}) \\ &< s + \varepsilon < r. \end{aligned}$$

Por otro lado, teniendo en cuenta las propiedades de la Proposición 2.2.11,

$$\begin{aligned} \mathcal{C} &= (\mathcal{D}_c\mathcal{C} - \mathcal{D}_c\mathcal{B}) + \mathcal{B} \\ &= (\mathcal{D}_c\mathcal{C} - \mathcal{D}_c\mathcal{B}) + [(\mathcal{D}_c\mathcal{B} - \mathcal{D}_c\mathcal{A}) + \mathcal{A}] \\ &= [(\mathcal{D}_c\mathcal{C} - \mathcal{D}_c\mathcal{B}) + (\mathcal{D}_c\mathcal{B} - \mathcal{D}_c\mathcal{A})] + \mathcal{A} \\ &= (\mathcal{D}_c\mathcal{C} - \mathcal{D}_c\mathcal{A}) + \mathcal{A}. \end{aligned}$$

Por lo tanto, $\mathcal{C} \in N_r(\mathcal{A})$ y así hemos probado que $N_\varepsilon(\mathcal{B}) \subseteq N_r(\mathcal{A})$. Utilizando estas tres propiedades, es conocido que existe una única topología en \mathcal{T} tal que $\beta_{\mathcal{A}} = \{N_{1/n}(\mathcal{A})\}_{n \in \mathbb{N}}$ es un sistema (numerable) de entornos de \mathcal{A} . Sólo queda demostrar que define una topología Hausdorff.

Sea $\mathcal{A}, \mathcal{B} \in \mathcal{T}$ tal que $\mathcal{A} \neq \mathcal{B}$. Distinguiamos dos casos. En primer lugar, supongamos que $\mathcal{D}_c\mathcal{A} \neq \mathcal{D}_c\mathcal{B}$ y sea $r = q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B})/4 > 0$. Para probar que

$N_r(\mathcal{A}) \cap N_r(\mathcal{B}) = \emptyset$, supongamos que $\mathcal{C} \in N_r(\mathcal{A}) \cap N_r(\mathcal{B})$. En este caso,

$$\begin{aligned} 4r &= q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{B}) \leq q_1\phi_1(\mathcal{D}_c\mathcal{A}, \mathcal{D}_c\mathcal{C}) + q_1\phi_1(\mathcal{D}_c\mathcal{C}, \mathcal{D}_c\mathcal{B}) \\ &< r + r = 2r, \end{aligned}$$

lo cual es absurdo. Por lo tanto, $N_r(\mathcal{A}) \cap N_r(\mathcal{B}) = \emptyset$.

Ahora supongamos que $\mathcal{D}_c\mathcal{A} = \mathcal{D}_c\mathcal{B}$. Como $\mathcal{C} \in N_r(\mathcal{A}) \cap N_r(\mathcal{B})$, entonces $\mathcal{C} = (\mathcal{D}_c\mathcal{C} - \mathcal{D}_c\mathcal{A}) + \mathcal{A}$ y $\mathcal{C} = (\mathcal{D}_c\mathcal{C} - \mathcal{D}_c\mathcal{B}) + \mathcal{B}$. Por lo tanto, podemos probar que $\mathcal{A} = (\mathcal{D}_c\mathcal{A} - \mathcal{D}_c\mathcal{C}) + \mathcal{C}$ y $\mathcal{B} = (\mathcal{D}_c\mathcal{B} - \mathcal{D}_c\mathcal{C}) + \mathcal{C}$ (utilizando la tercera propiedad de la Proposición 2.2.11). Como $\mathcal{D}_c\mathcal{A} = \mathcal{D}_c\mathcal{B}$, se deduce que $\mathcal{A} = \mathcal{B}$, pero esto es imposible. Por lo tanto, $N_r(\mathcal{A}) \cap N_r(\mathcal{B}) = \emptyset$ y, en cualquier caso, la topología generada por los sistemas de entornos dados es Hausdorff. ■

3.4.2. Estudio comparativo con los resultados de otros autores

Adabitabar Firozja *et al.* [1] realizaron un estudio de medidas de similitud basado en distancias intervalares y lo compararon con cuatro métodos que ya habían sido estudiados por otros investigadores (véase [38, 10, 11, 27]) utilizando los mismos datos. Los resultados de la comparación se muestran en [1, Tabla 2].

Las Tablas 3.2 y 3.3 muestran los conjuntos de números difusos (triangulares o trapezoidales) que se emplearon y las distancias que los distintos autores obtuvieron entre dichos números difusos. En concreto, la Tabla 3.2 muestra medidas de similitud reales entre los números difusos propuestos obtenidas por Lee [38], Chen [10], y Chen y Chen [11]. De la misma forma, la Tabla 3.3 presenta las medidas de similitud obtenidas por Guha y Chakraborty [27], y por Firozja y otros [1]. Además, en la Tabla 3.3, mostramos las medidas de similitud que se

Tabla 3.2: Conjuntos de números de difusos $\{\mathcal{A}, \mathcal{B}\}$ y medidas de similitud obtenidas por algunos autores.

Conjunto	\mathcal{A}	\mathcal{B}		
Conjunto 1	(0.1/0.2/0.3/0.4)	(0.1/0.25/0.4)		
Conjunto 2	(0.1/0.2/0.3/0.4)	(0.4/0.55/0.7)		
Conjunto 3	(0.1/0.2/0.3/0.4)	(0.4/0.5/0.6/0.7)		
Conjunto 4	(0.3/0.3/0.3)	(0.3/0.3/0.3)		
Conjunto 5	(0.2/0.2/0.2)	(0.3/0.3/0.3)		
Conjunto 6	(0.1/0.2/0.3)	(0.3/0.3/0.3)		
Conjunto 7	(0.1/0.2/0.3)	(0.2/0.3/0.4)		
Conjunto	Lee [38]	Chen [10]	Chen and Chen [11]	
Conjunto 1	0.9167	0.975	0.8357	
Conjunto 2	0.5	0.7	0.42	
Conjunto 3	0.5	0.7	0.49	
Conjunto 4	#	1	1	
Conjunto 5	0	0.9	0.9	
Conjunto 6	0.5	0.9	0.54	
Conjunto 7	0.6667	0.9	0.81	

obtienen empleando nuestra metodología cuando $\mathcal{D} = \mathcal{D}_c$, $h_1(\alpha) = h_2(\alpha) = 1 - \alpha$ y $\phi_i(x, y) = \psi_2(x, y) = |x - y|^p$, donde $p \in \{1, 0.5\}$.

Obsérvese que, por un lado, nuestro método produce un número difuso crisp cuando \mathcal{B} es una traslación de \mathcal{A} . Y, por otro lado, también se observa que nuestra metodología extiende el caso real al caso difuso. Pensamos, pues, que ambas son propiedades naturales desde el punto de vista intuitivo y lógico, lo que justifica el interés por las distancias de similitud que presentamos.

Tabla 3.3: Medidas difusas obtenidas por otros investigadores y medidas obtenidas a partir de nuestra metodología utilizando $p = 1$ y $p = 0,5$.

Conjunto	Guha y Chakraborty [27]	Firozja y otros [1]
Conjunto 1	(0.825/0.95/1)	[0.952, 1]
Conjunto 2	(0.525/0.65/0.75/0.877)	[0.741, 0.815]
Conjunto 3	(0.5/0.6/0.8/0.9)	[0.77, 0.77]
Conjunto 4	(1/1/1)	[1, 1]
Conjunto 5	(0.9/0.9/0.9)	[0.9, 0.9]
Conjunto 6	(0.85/0.9/0.95)	[0.833, 1]
Conjunto 7	(0.8/0.9/1)	[0.91, 0.91]
Conjunto	$p = 1$	$p = 0.5$
Conjunto 1	(-0.1/ -0.05/0.05/0.1)	(-0.447/ -0.224/0.224/0.447)
Conjunto 2	(0.2/0.25/0.35/0.4)	(0.1/0.324/0.771/0.995)
Conjunto 3	0.3	0.548
Conjunto 4	0	0
Conjunto 5	0.1	0.316
Conjunto 6	(0/0.1/0.2)	(0/0.316/0.632)
Conjunto 7	0.1	0.316

3.4.3. Ranking

En esta sección presentamos un ranking sobre \mathcal{F} (no sólo sobre \mathcal{T}) inspirado por la medida de similitud D que se basa en comparar dos números difusos mejor que dar como resultado un único número real. Sea $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$ una defusificación, y sea $d = [\underline{d}, \bar{d}] : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_0^+$ una medida de similitud con respecto a un orden parcial \sqsubseteq sobre \mathcal{I} . Sean $\mathcal{A}, \mathcal{B} \in \mathcal{F}$ dos números difusos diferentes. Si $\mathcal{A} \preceq \mathcal{B}$ (respectivamente, $\mathcal{B} \preceq \mathcal{A}$) con respecto a $(\mathcal{D}, d, \sqsubseteq)$, entonces escribiremos $\mathcal{A} \prec \mathcal{B}$

(respectivamente, $\mathcal{B} \prec \mathcal{A}$). Supongamos que $\mathcal{A} \preceq \mathcal{B}$ y que $\mathcal{B} \preceq \mathcal{A}$ con respecto a $(\mathcal{D}, d, \sqsubseteq)$. En este caso, aplicaremos el siguiente algoritmo.

- Si $\mathcal{DA} < \mathcal{DB}$ (respectivamente, $\mathcal{DB} < \mathcal{DA}$), escribiremos $\mathcal{A} \prec \mathcal{B}$ (respectivamente, $\mathcal{B} \prec \mathcal{A}$).
- Si $\mathcal{DA} = \mathcal{DB}$ y $\text{spr } \mathcal{A} < \text{spr } \mathcal{B}$ (respectivamente, $\text{spr } \mathcal{B} < \text{spr } \mathcal{A}$), entonces escribiremos $\mathcal{A} \prec \mathcal{B}$ (respectivamente, $\mathcal{B} \prec \mathcal{A}$).
- Si $\mathcal{DA} = \mathcal{DB}$, $\text{spr } \mathcal{A} = \text{spr } \mathcal{B}$ y $\underline{d}(\ker \mathcal{A}, \text{supp } \mathcal{A}) > \underline{d}(\ker \mathcal{B}, \text{supp } \mathcal{B})$ (respectivamente, $\underline{d}(\ker \mathcal{A}, \text{supp } \mathcal{A}) < \underline{d}(\ker \mathcal{B}, \text{supp } \mathcal{B})$), entonces escribiremos $\mathcal{A} \prec \mathcal{B}$ (respectivamente, $\mathcal{B} \prec \mathcal{A}$).
- Finalmente, si suponemos que $\mathcal{DA} = \mathcal{DB}$, $\text{spr } \mathcal{A} = \text{spr } \mathcal{B}$, $\underline{d}(\ker \mathcal{A}, \text{supp } \mathcal{A}) = \underline{d}(\ker \mathcal{B}, \text{supp } \mathcal{B})$ y $\bar{d}(\ker \mathcal{A}, \text{supp } \mathcal{A}) < \bar{d}(\ker \mathcal{B}, \text{supp } \mathcal{B})$ (respectivamente, $\bar{d}(\ker \mathcal{A}, \text{supp } \mathcal{A}) > \bar{d}(\ker \mathcal{B}, \text{supp } \mathcal{B})$), entonces escribiremos $\mathcal{A} \prec \mathcal{B}$ (respectivamente, $\mathcal{B} \prec \mathcal{A}$).

A continuación, vamos a comparar nuestra metodología con otros procesos de ranking. En [45], los autores describieron un nuevo método para calcular el ranking entre números difusos, y compararon sus algoritmos con metodologías que se habían propuesto anteriormente para ello (véase [45, Tabla 5]). Datos similares ya se consideraron en [61]. Si usamos nuestra metodología considerando el área como la defusificación, obtenemos los siguientes resultados, que están de acuerdo con los resultados obtenidos por la mayoría de los métodos de ranking previos.

La Tabla 3.4 muestra los resultados del ranking obtenido utilizando la metodología propuesta. Esta tabla pone de manifiesto que nuestra definición de medida de similitud permite la comparación de números difusos que pertenecen a clases distintas.

Tabla 3.4: Ranking obtenido utilizando nuestra metodología.

	\mathcal{A}_1	\mathcal{A}_2	\mathcal{A}_3
Conjunto 1	(0.4/0.5/1)	(0.4/0.7/1)	(0.4/0.9/1)
Conjunto 2	(0.3/0.4/0.7/0.9)	(0.3/0.7/0.9)	(0.5/0.7/0.9)
Conjunto 3	(0.3/0.5/0.7)	(0.3/0.5/0.8/0.9)	(0.3/0.5/0.9)
Conjunto 4	(0/0.4/0.7/0.8)	(0.2/0.5/0.9)	(0.1/0.6/0.8)
Ordenación			
Conjunto 1	$\mathcal{A}_1 \prec \mathcal{A}_2 \prec \mathcal{A}_3$		
Conjunto 2	$\mathcal{A}_1 \prec \mathcal{A}_2 \prec \mathcal{A}_3$		
Conjunto 3	$\mathcal{A}_1 \prec \mathcal{A}_3 \prec \mathcal{A}_2$		
Conjunto 4	$\mathcal{A}_1 \prec \mathcal{A}_2 \prec \mathcal{A}_3$		

CAPÍTULO 4

Métodos de regresión difusa basados en distancias difusas

El análisis de regresión es una poderosa herramienta estadística que tiene muchas aplicaciones en diferentes áreas. El problema del análisis de regresión en un entorno difuso ha sido tratado en el pasado desde diferentes puntos de vista, teniendo en cuenta una amplia variedad de procedimientos y de datos de entrada y de salida (reales o difusos). El objetivo principal de este capítulo es describir una nueva metodología para resolver problemas de regresión difusos considerando las dos nociones fundamentales introducidas en el capítulo anterior: métricas (o semimétricas) difusas y el orden parcial sobre el conjunto de números difusos trapezoidales.

4.1. Un primer acercamiento al problema de regresión difusa

Cuando afrontamos un problema de regresión difusa, surgen, de manera natural, cinco cuestiones íntimamente relacionadas con la incertidumbre inherente

al modelo, y que han sido tratadas con mayor o menor profundidad en el pasado.

1. En primer lugar, aunque el problema del análisis de regresión en un entorno difuso ha sido estudiado por diferentes autores desde diversos puntos de vista y teniendo en cuenta una amplia variedad de datos (tanto de entrada como de salida), nos damos cuenta de que, en general, la mayoría de los trabajos de investigación consideran distancias reales. Esto conlleva un conflicto entre la solución del problema de regresión difusa y la interpretación de los datos difusos como distribuciones de posibilidad (véase, por ejemplo, Näther [44], Coppi *et al.* [15], Roldán *et al.* [53] y las referencias que se encuentran dentro de los mismos, ya que proporcionan revisiones extensas de los principales enfoques del problema de regresión difusa), es decir, afrontamos un problema difuso utilizando técnicas reales.
2. En segundo lugar, si las distancias reales no son coherentes en este entorno, debemos plantearnos el uso de distancias difusas para llevar a cabo el proceso de regresión. En este sentido, debido a sus posibles aplicaciones en diversas áreas, la noción de *métrica* juega un papel clave en muchos campos de estudio que interpretan la medida de la distancia entre dos puntos como la diferencia entre ellos. Tradicionalmente, la distancia entre dos puntos ha sido siempre un número real. Incluso en un contexto difuso, la distancia entre dos *números difusos* suele interpretarse como un número real considerando, por ejemplo, el área entre ellos (véase [12, 53]). Sin embargo, esta concepción no capta completamente la imprecisión e incertidumbre del proceso y, en consecuencia, no es consistente con factores como la vaguedad y la ambigüedad que afectan al comportamiento del fenómeno estudiado en el entorno difuso. Por lo tanto, es más razonable utilizar un número difuso, en lugar de un número real, para medir la distancia entre dos números difusos.

3. En estrecha relación con el problema anterior surge la necesidad de introducir un orden parcial en el conjunto de los datos difusos que nos permita encontrar el modelo que alcanza el menor error difuso.
4. Cuando consideramos variables aleatorias difusas en el análisis de regresión, este problema se vuelve, en general, muy complejo y podemos encontrar grandes dificultades en la búsqueda de soluciones óptimas, especialmente en el caso de problemas no lineales. Esto ha llevado a la mayoría de los/as investigadores/as a considerar este problema únicamente en situaciones muy particulares. Para simplificar el modelo, se han venido considerando hasta ahora números difusos con formas geométricas muy simples, es decir, los datos se han restringido a intervalos numéricos o, como mucho, a números difusos triangulares y, además, el modelo se limita al caso lineal.
5. Por último, teniendo en cuenta que un estudio empírico puede considerar, al mismo tiempo, variables reales y variables difusas, una metodología de regresión difusa ha de ser capaz de tratar, de una manera unificada, ambos casos. Las metodologías que sólo consideran números difusos no reales (por ejemplo, los números difusos triangulares o trapezoidales) no pueden dar una respuesta adecuada a este problema cuando alguna de las variables que interviene en el problema no es difusa sino real. Nuestra metodología proporciona un procedimiento para cubrir ambas posibilidades. En este sentido, desde nuestro punto de vista, las métricas difusas deben extender el caso de las métricas reales.

Las tres primeras cuestiones planteadas quedaron resueltas en el capítulo anterior. Por ello, utilizando el orden parcial difuso y la familia de medidas de similitud entre números difusos introducidas en dicho capítulo, el objetivo

consiste ahora en desarrollar una nueva metodología que tenga en cuenta las siguientes características:

- debe permitir obtener modelos de regresión difusos tanto lineales como no lineales;
- debe utilizar un proceso de estimación que, en general, pueda considerarse fácil de aplicar en la práctica (a diferencia de otros métodos cuya aplicación práctica resulte ser más compleja); y
- debe considerar una amplia variedad de datos difusos, que sean apropiados para manejar la incertidumbre presente en los estudios de la vida real (y que, por lo tanto, no se limite a intervalos o a números difusos triangulares).

Los pasos que vamos a seguir en las siguientes secciones son los siguientes. En primer lugar, vamos a establecer el modelo de regresión con variables explicativas reales y una variable respuesta trapezoidal a partir de las medidas difusas de similitud entre los números difusos trapezoidales introducidas en el capítulo anterior. A continuación obtendremos el error difuso total y la solución óptima considerando el orden parcial \preceq . Para resolver el procedimiento de estimación, distinguiremos tres casos, según si las amplitudes dependen (o no) de variables explicativas reales, o de si se considera como entrada un vector aleatorio difuso. Desarrollaremos nuestra metodología de la manera más general posible (con objeto de poder afrontar problemas de regresión muy diversos) si bien, por simplicidad, en los ejemplos prácticos que presentaremos en el siguiente capítulo, utilizaremos casos muy concretos (como la defusificación \mathcal{D}_c).

En la Figura 4.1 resumimos el proceso que vamos a desarrollar en la siguiente sección.

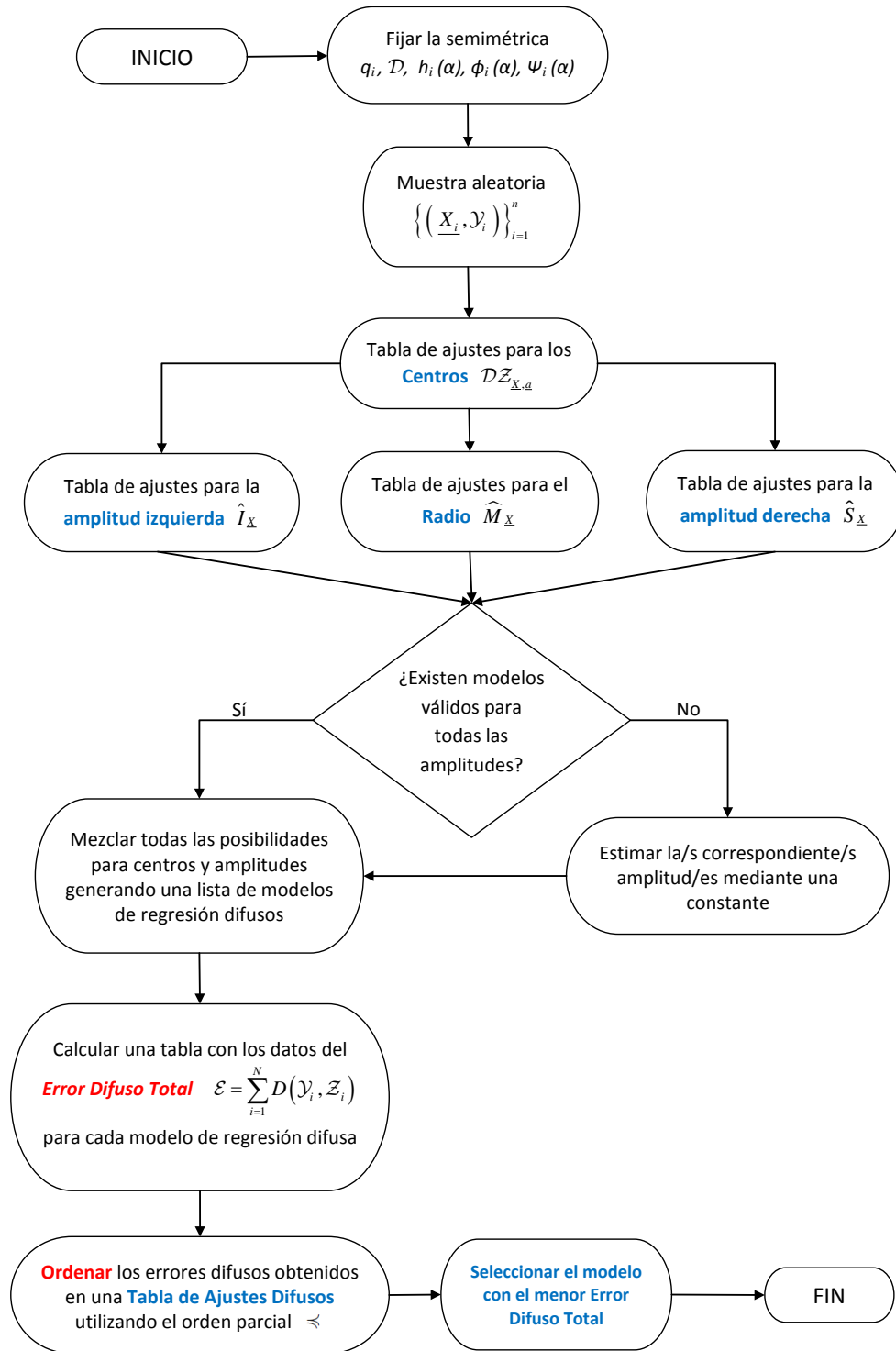


Figura 4.1: Diagrama de flujo que resume el método de regresión difusa.

4.2. El planteamiento del problema de regresión difusa en el caso trapezoidal

El problema que afrontaremos en este capítulo puede ser planteado de la forma que sigue. Sean \underline{X} e \mathcal{Y} dos variables aleatorias donde $\underline{X} = (X_1, \dots, X_N)'$ es un vector aleatorio real e \mathcal{Y} es una variable aleatoria trapezoidal¹. De entrada, podemos decir que si la variable respuesta \mathcal{Y} no fuese trapezoidal, siempre es posible considerar un operador de aproximación $\mathcal{O} : \mathcal{F} \rightarrow \mathcal{T}$ que transforme cualquier número difuso en un número difuso trapezoidal \mathcal{Y}' y que sea, en algún sentido, el más cercano al número difuso original (por ejemplo, puede considerarse, dentro del conjunto de todos los números difusos trapezoidales, el que más se aproxima con un intervalo esperado idéntico al del número difuso original, véase [26]).

Supongamos que estamos interesados en explicar el comportamiento de la variable \mathcal{Y} a partir de los valores de la variable \underline{X} . En este caso, el modelo de regresión que vamos a considerar puede formalizarse como:

$$\mathcal{Y} = \text{Tra}(Y_{\underline{X}, \underline{a}_c}^c + \varepsilon^c, M_{\underline{X}, \underline{a}_m} + \varepsilon^m, I_{\underline{X}, \underline{a}_\ell} + \varepsilon^\ell, S_{\underline{X}, \underline{a}_s} + \varepsilon^s), \quad (4.1)$$

donde $\varepsilon^c, \varepsilon^m, \varepsilon^\ell$ y ε^s son los residuos (es decir, variables aleatorias reales tales que $E[\varepsilon^c | \underline{X}] = E[\varepsilon^m | \underline{X}] = E[\varepsilon^\ell | \underline{X}] = E[\varepsilon^s | \underline{X}] = 0$ y cuyas varianzas son finitas) y $\underline{a}_c = (a_{c1}, \dots, a_{cN})'$, $\underline{a}_m = (a_{m1}, \dots, a_{mN})'$, $\underline{a}_\ell = (a_{\ell1}, \dots, a_{\ell N})'$ y $\underline{a}_s = (a_{s1}, \dots, a_{sN})'$ son los $(N \times 1)$ -vectores de los parámetros asociados al vector aleatorio \underline{X} . Por lo tanto, la esperanza condicionada, es decir, la función de regresión de la población, es:

$$E[\mathcal{Y} | \underline{X}] = \text{Tra}(Y_{\underline{X}, \underline{a}_c}^c, M_{\underline{X}, \underline{a}_m}, I_{\underline{X}, \underline{a}_\ell}, S_{\underline{X}, \underline{a}_s}).$$

Para predecir los valores de la variable difusa \mathcal{Y} a partir del vector \underline{X} , nuestro

¹Más adelante, en la Subsección §4.3.3, nos plantearemos el caso en el que $\underline{X} = (\mathcal{X}_1, \dots, \mathcal{X}_N)'$ es un vector de variables aleatorias difusas.

objetivo será determinar una aplicación

$$\mathcal{Z} : \text{dom } \underline{X} \rightarrow \mathcal{T} \quad (4.2)$$

que modelice dicho comportamiento. Obsérvese que cualquier aplicación de la forma que estamos buscando (que pueda servir como solución del problema de regresión) debe poder expresarse utilizando la notación introducida en (2.4), de manera que podemos suponer que

$$\mathcal{Z}_{\underline{X}} = \text{Tra}(Y_{\underline{X}}^c, M_{\underline{X}}, I_{\underline{X}}, S_{\underline{X}}),$$

donde

$$Y^c : \text{dom } \underline{X} \rightarrow \mathbb{R} \quad \text{y} \quad M, I, S : \text{dom } \underline{X} \rightarrow \mathbb{R}_0^+$$

son funciones arbitrarias. Resolver el problema de regresión difusa será determinar las mejores funciones \widehat{Y}^c , \widehat{M} , \widehat{I} y \widehat{S} , dentro de una clase concreta de funciones dependientes de ciertos parámetros, para predecir el comportamiento de la variable \mathcal{Y} dependiendo de los valores que toma el vector \underline{X} . Es muy importante observar que la primera función (\widehat{Y}^c) modelizará el centro del número difuso, por lo que puede tomar cualquier valor real, pero las otras tres (\widehat{M} , \widehat{I} y \widehat{S}) describen el comportamiento de las amplitudes, por lo que sólo pueden tomar valores no negativos.

Para afrontar el problema de determinar las funciones \widehat{Y}^c , \widehat{M} , \widehat{I} y \widehat{S} que intervienen en la modelización, habrá que tener en cuenta tres factores.

- En primer lugar, debemos determinar una forma de medir la distancia que existe entre dos números difusos dados. Para ello, utilizaremos una medida de similitud D del tipo de las aplicaciones definidas por las ecuaciones (3.8)-(3.9). Como vimos en el capítulo anterior, esta aplicación $D : \mathcal{F} \times \mathcal{F} \rightarrow \mathcal{F}$ se interpreta como una forma de decidir cómo de parecidos o de diferentes son dos números difusos \mathcal{A} y \mathcal{B} en términos de otro número difuso $D(\mathcal{A}, \mathcal{B})$. En general, esta aplicación verifica propiedades muy similares a las de una auténtica métrica, y tiene la ventaja de

que, bajo ciertas suposiciones, podemos afirmar que se comporta, sobre el conjunto de números difusos trapezoidales, como una auténtica métrica $D : \mathcal{T} \times \mathcal{T} \rightarrow \mathcal{T}$ (recuérdese el Teorema 3.3.16). Es por ello que, desde el principio de la investigación, deben fijarse los elementos variables que intervienen en las expresiones (3.8)-(3.9), a saber, la defusificación $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$, las constantes no negativas $q_1, q_2 \geq 0$, las funciones (no crecientes y continuas por la izquierda sobre \mathbb{I}) $h_1, h_2 : \mathbb{I} \rightarrow [0, \infty)$, las pseudosemimétricas $\phi_1, \phi_3, \phi_4 : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}_0^+$ consideradas sobre \mathbb{R} y la pseudosemimétrica $\psi_2 : \mathbb{R}_0^+ \times \mathbb{R}_0^+ \rightarrow \mathbb{R}_0^+$ considerada sobre \mathbb{R}_0^+ . Todas estas variables dependerán de las necesidades de la investigación: por ejemplo, considerar una pseudosemimétrica de mínimos cuadrados $\phi_i(x, y) = (x - y)^2$ producirá una única solución, la cual es estable para pequeñas variaciones de las condiciones iniciales, pero una pseudosemimétrica de desviaciones absolutas $\psi_2(x, y) = |x - y|$, a pesar de no tener las ventajas anteriores, protege frente a valores atípicos.

- Una vez fijada la medida de similitud (o métrica) D , también es importante determinar, a priori, los parámetros de los que depende la relación binaria \preceq considerada en la Definición 3.3.1. Aunque, en general, la relación \preceq sólo es reflexiva y transitiva (lo que se conoce como un *preorden*), también es verdad que, bajo ciertas condiciones, se transforma en un orden parcial sobre \mathcal{T} (recuérdese el Lema 3.3.2). En este sentido y, dependiendo de las características de la investigación desarrollada, han de fijarse el orden parcial \sqsubseteq sobre \mathcal{I} y la distancia intervalar $d = [\underline{d}, \bar{d}] : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{I}_{0_{\mathcal{I}}, \sqsubseteq}^+$ sobre $(\mathcal{I}, 0_{\mathcal{I}}, \sqsubseteq, s_{\mathcal{I}})$ en el sentido de la Definición 3.1.2 (la defusificación $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$ ya se fijó en el punto anterior).
- Una vez fijados los aspectos anteriores, consideremos un experimento aleatorio en el que observamos la variable conjunta $(\underline{X}, \mathcal{Y})$ sobre n unidades estadísticas, es decir, supongamos que disponemos de una muestra alea-

toria

$$\{\underline{X}_i, \mathcal{Y}_i = \text{Tra}(Y_i, M_i, I_i, S_i)\}_{i=1}^n$$

obtenida a partir de $(\underline{X}, \mathcal{Y})$.

Utilizando los tres elementos anteriores, determinaremos el *error difuso total* del modelo (4.1) asociado a la muestra aleatoria simple $\{\underline{X}_i, \mathcal{Y}_i\}_{i=1}^n$, que viene determinado por la expresión

$$\mathcal{E} = \sum_{i=1}^n D(\mathcal{Y}_i, \underline{Z}_{\underline{X}_i}). \quad (4.3)$$

Como puede observarse, el valor de \mathcal{E} no es real, sino que es un número difuso. Su valor concreto dependerá de los vectores de parámetros \underline{a}_c , \underline{a}_m , \underline{a}_l y \underline{a}_s que se elijan para determinar las funciones Y^c , M , I y S dentro de ciertas clases concretas de funciones (polinómicas, exponenciales, logarítmicas, etc.). Si la medida de similitud D se ha tomado para que ésta sea una auténtica métrica sobre \mathcal{T} , entonces el valor del error difuso total será un número difuso trapezoidal. De esta forma, el objetivo principal del problema de regresión difusa que nos planteamos será el de determinar una aplicación

$$\widehat{\mathcal{Z}} : \text{dom } \underline{X} \rightarrow \mathcal{T}, \quad \widehat{\mathcal{Z}}_{\underline{X}} = E[\mathcal{Y}|\underline{X}] = \text{Tra}(\widehat{Y}_{\underline{X}}^c, \widehat{M}_{\underline{X}}, \widehat{I}_{\underline{X}}, \widehat{S}_{\underline{X}})$$

tal que el error difuso total (4.3) sea tan pequeño como sea posible con respecto a la relación binaria \preceq sobre \mathcal{T} (que, dependiendo de cómo se hayan elegido sus factores, puede ser un auténtico orden parcial sobre \mathcal{T}).

Con este objetivo en mente, debemos hacer las siguientes apreciaciones.

- ▷ En realidad, a lo largo de la presente memoria, nos centraremos en el caso difuso trapezoidal pues, en esta situación, podemos disponer de un orden parcial y de una auténtica métrica difusa. No obstante, otras investigaciones en las que los datos de entrada y/o salida no fuesen trapezoidales,

pueden ser afrontadas de la misma forma, utilizando una medida de similitud D (aunque no sea una métrica) y un preorden \preceq (aunque no sea un orden parcial). Dejamos este caso para investigaciones futuras, y nos centraremos en el caso trapezoidal, que es suficientemente general como para modelizar una enorme cantidad de experimentos difusos aleatorios.

- ▷ Una primera forma de afrontar el problema anterior podría ser la siguiente: teniendo en cuenta que el objetivo final es el de encontrar cuatro funciones $Y^c : \text{dom } \underline{X} \rightarrow \mathbb{R}$ y $M, I, S : \text{dom } \underline{X} \rightarrow \mathbb{R}_0^+$ que sirvan para establecer el modelo $\mathcal{Z}_{\underline{X}} = \text{Tra}(Y_{\underline{X}}^c, M_{\underline{X}}, I_{\underline{X}}, S_{\underline{X}})$, cabe la posibilidad de desarrollar cuatro procedimientos de regresión real (por mínimos cuadrados) que determinen el modelo difuso. Como resultado de estos procesos, elegiríamos para Y^c , M , I y S aquellas funciones que tuviesen un mayor coeficiente de determinación real, R^2 . Este procedimiento ya ha sido propuesto en el pasado y, como veremos más adelante, presenta algunas deficiencias importantes desde el punto de vista teórico (como la necesidad de no negatividad para las amplitudes M , I y S). Queremos advertir de que no va a ser ésta, en absoluto, nuestra metodología. Como detallaremos a continuación, nosotros también proponemos buscar varias funciones tanto para modelizar los centros como para las amplitudes (posiblemente, dependiendo de las necesidades de la investigación, se elija el método de mínimos cuadrados, pero otros métodos también serían admisibles). Sin embargo, la validez de dichos modelos no vendrá determinada por un coeficiente R^2 real, sino que habrá que combinar varios de estos modelos para determinar el error difuso total, que será un número difuso, dependiendo de una métrica (o medida de similitud) plenamente difusa. De esta forma, utilizando el orden parcial que ya introdujimos, trataremos de determinar la combinación más adecuada de modelos para los centros y las amplitudes que produzca el menor error difuso total. Así, nuestra metodología está completamente basada en nociones difusas y no reales: no se utilizan distancias reales sino

difusas, y no buscamos un modelo óptimo según un coeficiente de determinación real, sino empleando un orden parcial en el conjunto de números difusos. Como veremos, esta metodología no siempre lleva a que el modelo con el menor coeficiente R^2 real es el más apropiado. Es más, esta metodología suele producir modelos (en la mayoría de ocasiones, no lineales) más eficientes que los generados por los procedimientos, reales o difusos, descritos por otros autores/as.

▷ Como hemos visto anteriormente, para considerar la medida de similitud D y la relación binaria \preceq hay que fijar, desde el principio de la investigación, varias constantes, funciones y pseudosemimétricas reales de las que dependerán las características fundamentales de D y de \preceq . Esto dota al investigador/a de una enorme cantidad de posibilidades diferentes (lo cual es una de las ventajas de nuestra metodología). No obstante, ante elecciones particulares de todas estas variables, podemos observar (como hacemos en el siguiente teorema) que el archiconocido método de regresión por mínimos cuadrados puede verse como un caso concreto de todas las variables que intervienen en nuestro proceso de estimación.

Teorema 4.2.1 Si $\mathcal{D} = \mathcal{D}_c$, h_1 y h_2 son dos negaciones estándar, $q_0 = q = 1$ y $\phi_0(x, y) = \psi(x, y) = \phi_i(x, y) = \varphi_i(x, y) = |x - y|^2$ para cualesquiera x, y en sus respectivos dominios, entonces la solución del correspondiente problema difuso de regresión coincide con la solución obtenida por el método de mínimos cuadrados.

Este resultado nos anima a considerar el método de mínimos cuadrados, que ha sido ampliamente utilizado en la literatura, por diversas razones: (1) existe una gran cantidad de software, incluso gratuito, que puede emplearse; (2) permite obtener modelos lineales y no lineales; (3) para el caso de datos reales, está disponible en muchos paquetes estadísticos; (4) es fácil de usar y comprender; (5) en muchos casos, proporciona buenas aproximaciones; (6) tiene

ventajas en el cálculo; (7) es muy adecuado para el aprendizaje, etc. Nuestro propósito no es mejorar este método, sino utilizar este criterio para desarrollar una metodología difusa que sea más útil y potente. Por eso, utilizaremos este procedimiento en la mayoría de las aplicaciones que presentaremos en el capítulo siguiente.

4.3. Una nueva metodología de regresión difusa

En esta sección explicamos la metodología que proponemos para resolver el problema de regresión difusa. Supongamos que hemos fijado todos los elementos necesarios para determinar la medida de similitud D y la relación binaria \preceq . Utilizando el Lema 2.2.12 dentro del conjunto formado por todos los números trapezoidales \mathcal{T} , podemos descomponer la función \mathcal{Z} de la siguiente forma,

$$\mathcal{Z} = \mathcal{D}\mathcal{Z} + \mathcal{D}^*\mathcal{Z},$$

donde $\mathcal{D}\mathcal{Z}_{\underline{x}} \in \mathbb{R}$ es un número real y $\mathcal{D}^*\mathcal{Z}_{\underline{x}} \in \mathcal{T}$ es un número trapezoidal. Estas dos partes diferentes del modelo nos obligan a dividir nuestra metodología en dos partes.

- En primer lugar, consideremos la muestra aleatoria real $\{(\underline{X}_i, \mathcal{D}\mathcal{Y}_i)\}_{i=1}^n$. Al fijar la defusificación $\mathcal{D} : \mathcal{F} \rightarrow \mathbb{R}$, estamos admitiendo que el punto que más nos interesa de cada número difuso \mathcal{Y}_i es $\mathcal{D}\mathcal{Y}_i$. Por tanto, nos parece razonable utilizar la muestra real $\{(\underline{X}_i, \mathcal{D}\mathcal{Y}_i)\}_{i=1}^n$ para aplicar técnicas de regresión reales, como son el método de mínimos cuadrados (que es ampliamente conocido y utilizado) y el método de desviaciones absolutas (que presenta otras ventajas: por ejemplo, se comporta mejor frente a valores atípicos). Estas técnicas nos permiten encontrar algunas funciones que pueden servir para modelar el comportamiento de la función que determine los centros de la solución del problema. No obstante,

no debemos conformarnos con una única función. Utilizando cualquiera de los paquetes estadísticos disponibles y la semidistancia real que más interese al investigador/a (o que tenga más sentido en el contexto del problema), se debe generar una amplia variedad de formas funcionales con el objeto de modelizar la función $\widehat{Y}^c = \widehat{\mathcal{D}}\widehat{\mathcal{Z}}$. Dichos modelos serán ordenados en orden decreciente del valor R^2 en una tabla (a este respecto, téngase en cuenta los matices que introduciremos en la Observación 4.4.1) que se denominará *Tabla de ajustes*.

- Toda vez que se tiene una tabla con los diferentes modelos que pudieran ser válidos para estimar el centro $\widehat{Y}^c = \widehat{\mathcal{D}}\widehat{\mathcal{Z}}$ del modelo que buscamos, pasamos a buscar una función de regresión para estimar la parte difusa trapezoidal:

$$\widehat{\mathcal{D}^*\mathcal{Z}}_{\underline{x}} = \text{Tra}(0, \widehat{M}_{\underline{x}}, \widehat{I}_{\underline{x}}, \widehat{S}_{\underline{x}}).$$

Es decir, nos planteamos encontrar las funciones \widehat{M} , \widehat{I} y \widehat{S} que determinarán esta segunda parte del modelo. En este caso, hay una importante diferencia con el apartado anterior: estas funciones no pueden tomar valores negativos, pues representan a las amplitudes de la aplicación difusa que servirá como solución del problema. Para afrontar esta parte, tendremos que distinguir entre dos casos distintos según si $\mathcal{D}^*\mathcal{Z}$ depende, o no, de las variables explicativas que estemos considerando. En el primer caso, si hay dependencia estadística, buscaremos modelos adecuados para las amplitudes (véase la Subsección §4.3.1). Si no es así, proponemos ajustar una función constante (como se describirá en la Subsección §4.3.2).

De esta forma, la combinación de un modelo para el centro $\widehat{Y}^c = \widehat{\mathcal{D}}\widehat{\mathcal{Z}}$ junto con un modelo para la parte difusa $\widehat{\mathcal{D}^*\mathcal{Z}}_{\underline{x}} = \text{Tra}(0, \widehat{M}_{\underline{x}}, \widehat{I}_{\underline{x}}, \widehat{S}_{\underline{x}})$, siguiendo las consideraciones previas e intentado que el error difuso total (4.3) sea lo menor posible respecto del orden parcial \preceq , nos conducirá al modelo de regresión difusa

resultante mediante la expresión:

$$\widehat{\underline{Z}}_x = \widehat{\mathcal{D}}\underline{Z} + \widehat{\mathcal{D}^*}\underline{Z}_x = (\widehat{Z}_x^1 / \widehat{Z}_x^2 / \widehat{Z}_x^3 / \widehat{Z}_x^4) \in \mathcal{T}. \quad (4.4)$$

Para describir los dos casos que pueden plantearse, supongamos que, toda vez que se ha confeccionado una tabla de ajustes que ofrezca diferentes funciones para modelizar la función \widehat{Y}^c que determinará los centros, tratamos de buscar otras funciones $M, I, S : \text{dom } \underline{X} \rightarrow \mathbb{R}_0^+$ para modelizar las amplitudes. En tal caso, recurriremos a las muestras aleatorias reales $\{(X_i, M_i)\}_{i=1}^n$, $\{(X_i, I_i)\}_{i=1}^n$ y $\{(X_i, S_i)\}_{i=1}^n$. Aplicando un método similar al que empleamos para obtener la anterior tabla de ajustes, deberemos buscar nuevos modelos que puedan servir para M, I y S . Dado que ninguna de las tres funciones M, I y S presenta características diferentes de las otras dos, en las dos siguientes subsecciones describimos cómo determinar una de ellas (en concreto, M), aplicándose el mismo proceso a las otras dos (eso sí, cabe la posibilidad de que una de ellas sí dependa de las variables explicativas y otra no, por lo que se aplicarían los dos métodos diferentes de las dos siguientes subsecciones).

Cuando se encuentran modelos reales con un coeficiente R^2 cercano a 1 (es decir, modelos *significativos*), podemos afirmar que las variables explicadas dependen del vector de variables explicativas $\underline{X} = (X_1, \dots, X_N)'$. Por contra, coeficientes R^2 cercanos a cero (modelos *no significativos*) nos impiden establecer tal afirmación, por lo que el método que emplearemos será diferente.

4.3.1. El procedimiento de estimación cuando las amplitudes dependen de las variables explicativas

Para determinar la función M del modelo, utilizaremos la muestra aleatoria real $\{(X_i, M_i)\}_{i=1}^n$ y, dependiendo de las pseudosemimétricas que se empleen en D , buscaremos diferentes formas funcionales para M . En este caso, supongamos que se han obtenido algunas posibilidades para M con coeficientes R^2 cercanos

a 1. En tal caso, volveremos a confeccionar una tabla de posibles ajustes para M , ordenándolos de forma decreciente según su coeficiente R^2 .

El software que normalmente se emplea para desarrollar estos procesos de regresión reales no es capaz de asegurar, a priori, que las diferentes opciones para la función M tomarán exclusivamente valores no negativos. A este respecto, varios autores han presentado metodologías difusas para tratar de resolver este problema. Nosotros no somos partidarios de establecer condiciones a priori para conseguir que las funciones que aquí se obtengan estén obligadas a tomar valores no negativos, porque dichas condiciones suelen llevar a modelos no significativos, lo que invalida el proceso de regresión. Todo al contrario, la metodología que aquí se propone está basada en conseguir cuantos más modelos significativos mejor y, a posteriori, se estudiará su signo dentro del dominio donde la variable explicativa \underline{X} tome valores (fuera de este dominio, no tiene sentido estudiar su signo). Por lo tanto, entre las diferentes funciones que podríamos calcular, debemos de ser cuidadosos de elegir los modelos que se ajustan mejor en el sentido de la Observación 4.4.1 de entre aquéllos que sólo tomen valores no negativos y estén bien definidos en el conjunto de todos los puntos para los que estamos interesados en obtener predicciones. Este proceso se ilustra en la Sección §5.2 del siguiente capítulo.

4.3.2. El procedimiento de estimación cuando alguno de las amplitudes no depende de las variables explicativas

Existen algunos casos en los que las amplitudes no dependen de las variables explicativas, o en los que el método descrito anteriormente no proporciona buenos resultados, ya que no da lugar a un modelo significativo para alguna de las amplitudes, o porque los modelos más significativos toman valores negativos en el dominio de \underline{X} . En estos casos, proponemos ajustar una constante no

negativa considerando una función de error como la que se describe a continuación. Hacemos hincapié en que este problema no ha sido considerado por otros investigadores utilizando una semimétrica difusa general D , como aquí hacemos.

A lo largo de la presente subsección, supongamos que para una o varias amplitudes no ha sido posible ajustar un modelo significativo, y que estamos interesados en estimar una constante. Para elegir dicha constante, supondremos, como en la Observación 3.3.12, que se han elegido

$$\begin{aligned} h_1(\alpha) = h_2(\alpha) = 1 - \alpha & \text{ para cada } \alpha \in \mathbb{I}, \\ \left. \begin{aligned} \phi_1(x, y) = q_1 |x^{k_1} - y^{k_1}|^{p_1} \\ \phi_3(x, y) = q_3 |x^{k_3} - y^{k_3}|^{p_3} \\ \phi_4(x, y) = q_4 |x^{k_4} - y^{k_4}|^{p_4} \end{aligned} \right\} & \text{ para cada } x, y \in \mathbb{R}, \\ \psi_2(x, y) = q_2 |x^{k_2} - y^{k_2}|^{p_2} & \text{ para cada } x, y \in \mathbb{R}_0^+, \end{aligned}$$

en las expresiones (3.8)-(3.9), donde $q_i, k_i, p_i > 0$ para cada $i \in \{1, 2, 3, 4\}$. En tal caso, como vimos en (3.11), la métrica D viene dada por

$$\begin{aligned} D(\mathcal{A}, \mathcal{B}) = \text{Tra} \left(q_1 \left| (\mathcal{D}\mathcal{A})^{k_1} - (\mathcal{D}\mathcal{B})^{k_1} \right|^{p_1}, q_2 \left| (A^m)^{k_2} - (B^m)^{k_2} \right|^{p_2}, \right. \\ \left. q_3 \left| (A^\ell)^{k_3} - (B^\ell)^{k_3} \right|^{p_3}, q_4 \left| (A^s)^{k_4} - (B^s)^{k_4} \right|^{p_4} \right) \quad (4.5) \end{aligned}$$

para cada $\mathcal{A} = \text{Tra}(A^c, A^m, A^\ell, A^s)$ y $\mathcal{B} = \text{Tra}(B^c, B^m, B^\ell, B^s)$, y el orden parcial \preceq es, según (3.13),

$$\mathcal{A} \preceq \mathcal{B} \Leftrightarrow \{ A^c \leq B^c, A^m \leq B^m, A^\ell \leq B^\ell, A^s \leq B^s \}.$$

Por consiguiente, si $\mathcal{Y}_i = \text{Tra}(Y_i, M_i, I_i, S_i)$ y $\mathcal{Z}_{\underline{X}_i} = \text{Tra}(Y_{\underline{X}_i}^c, M_{\underline{X}_i}, I_{\underline{X}_i}, S_{\underline{X}_i})$

para cada $i \in \{1, 2, \dots, n\}$, entonces el error difuso total es

$$\begin{aligned}
\mathcal{E} &= \sum_{i=1}^n D(\mathcal{Y}_i, \underline{\mathcal{Z}}_{X_i}) \\
&= \sum_{i=1}^n \text{Tra} \left(q_1 \left| (\mathcal{D}\mathcal{Y}_i)^{k_1} - (\mathcal{D}\underline{\mathcal{Z}}_i)^{k_1} \right|^{p_1}, q_2 \left| M_i^{k_2} - M_{\underline{X}_i}^{k_2} \right|^{p_2}, \right. \\
&\quad \left. q_3 \left| I_i^{k_3} - I_{\underline{X}_i}^{k_3} \right|^{p_3}, q_4 \left| S_i^{k_4} - S_{\underline{X}_i}^{k_4} \right|^{p_4} \right) \\
&= \text{Tra} \left(q_1 \sum_{i=1}^n \left| (\mathcal{D}\mathcal{Y}_i)^{k_1} - (\mathcal{D}\underline{\mathcal{Z}}_i)^{k_1} \right|^{p_1}, q_2 \sum_{i=1}^n \left| M_i^{k_2} - M_{\underline{X}_i}^{k_2} \right|^{p_2}, \right. \\
&\quad \left. q_3 \sum_{i=1}^n \left| I_i^{k_3} - I_{\underline{X}_i}^{k_3} \right|^{p_3}, q_4 \sum_{i=1}^n \left| S_i^{k_4} - S_{\underline{X}_i}^{k_4} \right|^{p_4} \right).
\end{aligned}$$

Dado que minimizar el error difuso total equivale, en este caso, a minimizar cada componente, y dado que las funciones de minimización de cualquier amplitud son similares, es suficiente aprender a determinar el mínimo absoluto de una función error $e :]0, \infty[\rightarrow [0, \infty[$ como la siguiente:

$$\begin{aligned}
e(x) &= \sum_{i=1}^n |x^k - a_i^k|^p \\
&= \begin{cases} \sum_{i=1}^n (a_i^k - x^k)^p, & \text{si } 0 < x \leq a_1, \\ \sum_{i=1}^r (x^k - a_i^k)^p + \sum_{i=r+1}^n (a_i^k - x^k)^p, & \text{si } a_r < x \leq a_{r+1} \\ & (1 \leq r \leq n-1), \\ \sum_{i=1}^n (x^k - a_i^k)^p, & \text{si } x > a_n, \end{cases} \quad (4.6)
\end{aligned}$$

donde $p > 0$, $k > 0$ y $0 < a_1 < a_2 < \dots < a_n$ (siendo $n \geq 2$ porque $n = 1$ es un caso trivial). Nótese que p_1 y k_1 se denotan aquí, por simplicidad, por p y k , respectivamente. Claramente, e es una función continua en $]0, \infty[$ y de clase C^∞ en $]0, \infty[\setminus \{a_1, a_2, \dots, a_n\}$. Más aún, la función e es estrictamente decreciente en $]0, a_1[$ y estrictamente creciente en $]a_n, \infty[$. Entonces e tiene, al menos, un mínimo absoluto en $]0, \infty[$, que está en $[a_1, a_n]$. Esto prueba que el problema de regresión difusa siempre tiene una solución. Para encontrarla distinguimos entre si $p > 1$, $p = 1$ o $0 < p < 1$.

Caso $p > 1$. Supongamos $p > 1$ en (4.6). Teniendo en cuenta la descomposición en partes de \mathbf{e} , podemos probar que

$$\lim_{x \rightarrow a_i^+} \mathbf{e}'(x) - \lim_{x \rightarrow a_i^-} \mathbf{e}'(x) = 0 \quad \text{para cada } i \in \{1, 2, \dots, n\},$$

por lo que la función \mathbf{e} es de clase C^1 en $]0, \infty[$. Además, su derivada es

$$\mathbf{e}'(x) = \begin{cases} -p k x^{k-1} \sum_{i=1}^n (a_i^k - x^k)^{p-1}, & \text{si } 0 < x \leq a_1, \\ p k x^{k-1} \left[\sum_{i=1}^r (x^k - a_i^k)^{p-1} - \sum_{i=r+1}^n (a_i^k - x^k)^{p-1} \right], & \text{si } a_r < x \leq a_{r+1} \\ & (1 \leq r \leq n-1), \\ p k x^{k-1} \sum_{i=1}^n (x^k - a_i^k)^{p-1}, & \text{si } x > a_n. \end{cases}$$

Como $\mathbf{e}'(x) < 0$ si $x \in]0, a_1[$ y $\mathbf{e}'(x) > 0$ si $x \in [a_n, \infty[$, el teorema de Bolzano, aplicado a \mathbf{e}' , nos garantiza que existe un valor $x_0 \in]a_1, a_n[$ tal que $\mathbf{e}'(x_0) = 0$. Como $p > 1$, entonces $\mathbf{e}''(x) > 0$ para cualquier $x \in]0, \infty[\setminus \{a_1, a_2, \dots, a_n\}$, por lo que \mathbf{e}' es estrictamente creciente en $]0, \infty[$. Por tanto, \mathbf{e}' tiene un único cero, que claramente es un mínimo absoluto. Para encontrar este punto, tenemos que resolver la ecuación $\mathbf{e}'(x_0) = 0$ en $]a_1, a_n[$, que es equivalente al sistema:

$$\left\{ \sum_{i=1}^r (x_0^k - a_i^k)^{p-1} = \sum_{i=r+1}^n (a_i^k - x_0^k)^{p-1} \text{ en el intervalo }]a_r, a_{r+1}[\right\}_{r=1}^{r=n-1}.$$

Caso $p = 1$. Si $p = 1$ en (4.6), entonces tanto \mathbf{e} como su primera derivada, vienen dadas por:

$$\mathbf{e}(x) = \sum_{i=1}^n |x^k - a_i^k|,$$

$$\mathbf{e}'(x) = \begin{cases} -k n x^{k-1}, & \text{si } 0 < x < a_1, \\ (2r - n) k x^{k-1}, & \text{si } a_r < x < a_{r+1}, \\ k n x^{k-1}, & \text{si } x > a_n. \end{cases}$$

En cada intervalo $]0, a_1[,]a_1, a_2[, \dots,]a_{n-1}, a_n[,]a_n, +\infty[$ el signo de \mathbf{e}' depende de los siguientes números:

$$M_r = 2r - n, \text{ para cada } r \in \{0, 1, 2, \dots, n\}.$$

Entonces

$$-n = M_0 < M_1 < M_2 < \dots < M_{n-1} < M_n = n.$$

En los primeros intervalos, e' es negativo (por lo tanto e es estrictamente decreciente) y en el final de los intervalos e' es positiva (por lo que e es estrictamente creciente). Si $M_r \neq 0$ para todo r , entonces e tiene un único mínimo absoluto, que es la mediana del conjunto de los datos $\{a_1, a_2, \dots, a_n\}$. Si $M_r = 0$ para algún r , entonces e alcanza su mínimo en todos los puntos del intervalo $[a_r, a_{r+1}]$ ya que la mediana no está unívocamente determinada entre a_r y a_{r+1} (son los valores centrales de la distribución).

Caso $0 < p < 1$. Si $0 < p < 1$ en (4.6), no es difícil probar que, para cada $r \in \{1, 2, \dots, n\}$,

$$\lim_{x \rightarrow a_{r+1}^-} e'(x) = -\infty \quad \text{y} \quad \lim_{x \rightarrow a_r^+} e'(x) = +\infty.$$

Si existe algún $x_0 \in]a_r, a_{r+1}[$ tal que $e'(x_0) = 0$, entonces podemos encontrar un valor $\varepsilon_0 > 0$ verificando que $]x_0 - \varepsilon_0, x_0 + \varepsilon_0[\subseteq]a_r, a_{r+1}[$ y $e'(x_0 - \varepsilon) > 0 > e'(x_0 + \varepsilon)$ para cada $\varepsilon \in]0, \varepsilon_0[$. Entonces, e' solo puede tener un cero en el intervalo $]a_r, a_{r+1}[$, y es creciente antes del cero y decreciente después del cero. Por lo tanto, este cero sólo puede ser un máximo local, por lo que el mínimo de e en cada intervalo $[a_r, a_{r+1}]$ debe estar en $\{a_r, a_{r+1}\}$. De esta forma, el mínimo absoluto de la función e en el intervalo $[a_1, a_n]$ debe estar en el conjunto de nodos $\{a_1, a_2, \dots, a_n\}$.

Resumimos todos los casos considerados en el siguiente teorema que da el estimador \hat{M} para M (los estimadores de I y S se obtienen mediante expresiones similares).

Teorema 4.3.1 Sean \underline{X} una variable aleatoria, sea \mathcal{Y} una variable aleatoria difusa trapezoidal verificando la ecuación (4.1) y sea $\{\underline{X}_i, \mathcal{Y}_i\}_{i=1}^n$ una muestra aleatoria obtenida de $(\underline{X}, \mathcal{Y})$, siendo $\mathcal{Y}_i = \text{Tra}(Y_i, M_i, I_i, S_i)$ para cada $i \in \{1, 2, \dots, n\}$. Sea \hat{Y}_X^c un modelo ajustado obtenido a partir de los datos

$\{\underline{X}_i, Y_i\}_{i=1}^n$. Entonces el problema de regresión (en $M, I, S \geq 0$) para alguna medida de similitud D definida en (4.5):

$$\begin{cases} \mathcal{Z} = \mathcal{Z}_{M,I,S} : \text{dom } \underline{X} \rightarrow \mathcal{T}, & \mathcal{Z}_X = \text{Tra}(\hat{Y}_X^c, M, I, S), \quad X \in \text{dom } \underline{X}, \\ E_{I,M,S} = \sum_{i=1}^n D(\mathcal{Y}_i, \mathcal{Z}_{\underline{X}_i}) & \text{mínimo con respecto a } \preceq, \end{cases}$$

tiene solución. El estimador \hat{M} para M es el siguiente estimador (estimadores para I y S pueden obtenerse de forma similar):

a) Si $p > 1$, \hat{M} es la única solución positiva del sistema de ecuaciones:

$$\hat{M} \rightarrow \left\{ \begin{array}{l} \sum_{i=1}^r (x^k - M_i^k)^{p-1} = \sum_{i=r+1}^n (M_i^k - x^k)^{p-1} \\ \text{en el intervalo }]M_r, M_{r+1}] \quad \begin{matrix} r=n-1 \\ r=1 \end{matrix} \end{array} \right\}. \quad (4.7)$$

b) Si $p = 1$, \hat{M} es la mediana de las amplitudes ($\hat{M} \rightarrow \text{median } M_i$).

c) Si $0 < p < 1$, \hat{M} se obtiene buscando los siguientes valores:

$$\hat{M} \rightarrow M_j \quad \text{tal que} \quad \left\{ \sum_{i=1}^n |M_j^k - M_i^k|^p \right\}_{j=1}^n \quad \text{es mínima.} \quad (4.8)$$

Además, \hat{M} está únicamente determinada en los siguientes casos:

- ▷ $p > 1$, o
- ▷ $p = 1$ y la mediana de las correspondientes amplitudes es única, o
- ▷ $0 < p < 1$ y los correspondientes valores en (4.8) tienen un único mínimo absoluto.

En el caso $p = 2$ (que es muy interesante porque lleva al método de mínimos cuadrado), el sistema (4.7) tiene una única solución:

$$\hat{M} = \sqrt[k]{\frac{M_1^k + M_2^k + \dots + M_n^k}{n}}.$$

En el caso $p = 3$, el sistema (4.7) es equivalente al sistema bicuadrático:

$$\left\{ \begin{array}{l} (2r - n)x^{2k} - 2x^k \left(\sum_{i=1}^r M_i^k - \sum_{i=r+1}^n M_i^k \right) \\ \quad + \left(\sum_{i=1}^r M_i^{2k} - \sum_{i=r+1}^n M_i^{2k} \right) = 0 \\ \text{en el intervalo }]M_r, M_{r+1}] \end{array} \right\}_{r=1}^{r=n-1} .$$

En situaciones prácticas, \widehat{M} , \widehat{I} y \widehat{S} pueden calcularse fácilmente como la solución de un problema de minimización, directamente programando y minimizando la función distancia.

4.3.3. Una extensión de la metodología previa

Nuestro modelo de regresión difusa se puede generalizar para el caso en el que las variables explicativas también son difusas, es decir, al caso en el que $\mathcal{X} = (\mathcal{X}_1, \dots, \mathcal{X}_N)'$ es un vector de variables difusas trapezoidales e \mathcal{Y} es una variable difusa trapezoidal. En base a una muestra aleatoria $\{\mathcal{X}_{1i}, \dots, \mathcal{X}_{Ni}, \mathcal{Y}_i\}_{i=1}^n$, si estamos interesados en analizar la relación entre \mathcal{Y} y $\mathcal{X}_1, \dots, \mathcal{X}_N$, entonces, este problema se puede reducir a los casos anteriores considerando como variable de entrada la vector aleatorio real

$$\underline{X}' = (X_1^1, X_1^2, X_1^3, X_1^4, X_2^1, \dots, X_{N-1}^4, X_N^1, X_N^2, X_N^3, X_N^4)' .$$

De hecho, si $\mathcal{X}_1, \dots, \mathcal{X}_N$ no son trapezoidales, también podemos considerar un operador de aproximación que lo transforme de manera adecuada.

4.4. Medidas de bondad de ajuste

Después de describir los procedimientos que proponemos para determinar las funciones que modelizarán el experimento aleatorio difuso, cabe plantearse una medida de bondad de ajuste que sea útil para evaluar nuestra metodología.

Dado que el error total que nosotros calcularemos es un número difuso y no un número real, nuestra metodología no es comparable con ningún otro procedimiento introducido hasta la fecha. No obstante, a falta de realizar estudios posteriores sobre esta cuestión, para poder mostrar la eficiencia de nuestro método, vamos a calcular, en cada uno de los ejemplos que presentemos, las medidas de bondad de ajuste que se emplean en el contexto real, aunque trasladadas al ambiente difuso. De esta forma veremos que, utilizando los estimadores clásicos de bondad de ajuste que surgen en el contexto real, nuestros modelos mejoran los existentes ya que conducen a un error menor en el sentido que comentamos en la siguiente observación.

Hemos visto que, aplicando la metodología anterior, obtendremos diferentes modelos de regresión difusa, que hemos denotado como

$$\mathcal{Z}_{\underline{x}} = \widehat{\mathcal{D}}_c \mathcal{Z}_{\underline{x}} + \widehat{\mathcal{D}}_c^* \mathcal{Z}_{\underline{x}} = (\mathcal{Z}_{\underline{x}}^1 / \mathcal{Z}_{\underline{x}}^2 / \mathcal{Z}_{\underline{x}}^3 / \mathcal{Z}_{\underline{x}}^4) \in \mathcal{T}.$$

Para evaluar la bondad de ajuste de los diferentes modelos, algunos/as autores/as han considerado principalmente las estimaciones numéricas dadas por los siguientes estadísticos.

- Por un lado, podemos calcular la *suma total de cuadrados debida al error* (véase [12, 53]), que denotamos por SSE , y que viene dada por:

$$SSE = \sum_{j=1}^4 SSE_j,$$

donde los valores $\{SSE_j\}_{j=1}^4$ se obtienen mediante

$$SSE_j = \sum_{i=1}^n \left(\mathcal{Y}_i^j - \mathcal{Z}_{\underline{x}_i}^j \right)^2, \quad 1 \leq j \leq 4,$$

siendo $\mathcal{Y}_i = (\mathcal{Y}_i^1 / \mathcal{Y}_i^2 / \mathcal{Y}_i^3 / \mathcal{Y}_i^4)$ para cada $i \in \{1, 2, \dots, n\}$. Algunos autores (véase [12]) también utilizaron como medida comparativa la *suma del error total*

$$\sum_{j=1}^4 E_j, \quad \text{donde } E_j = \frac{SSE_j}{4} \quad \text{para cada } j \in \{1, 2, 3, 4\}.$$

- Por otro lado, otros autores (véase, por ejemplo, [15]) consideraron un coeficiente R -cuadrado (o ajustado) dado, utilizando nuestra medida de distancia, por

$$\tilde{R}^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{j=1}^4 \sum_{i=1}^n (\mathcal{Y}_i^j - \mathcal{Z}_{\underline{x}_i}^j)^2}{\sum_{j=1}^4 \sum_{i=1}^n (\mathcal{Y}_i^j - \overline{\mathcal{Y}_i^j})^2},$$

donde

$$\overline{\mathcal{Y}_i^j} = \frac{1}{n} \sum_{i=1}^n \mathcal{Y}_i^j$$

es la media de los valores observados.

Observación 4.4.1 *Los modelos de regresión simples involucran una única variable de respuesta dependiendo de una única variable predictora. En este caso, los paquetes estadísticos ajustan una gran variedad de formas funcionales y calculan diferentes medidas estadísticas de la bondad de ajuste para modelos paramétricos: (1) la suma de cuadrados debida al error (SSE) y (2) R -cuadrado o R -cuadrado ajustado. Normalmente, los modelos se disponen en orden decreciente del coeficiente R -cuadrado. Recordemos que los valores del coeficiente R -cuadrado oscilan de 0 a 100%. Un valor R -cuadrado alto indicará un modelo útil mientras que un bajo R -cuadrado significa que el modelo correspondiente no debe ser tenido en cuenta por el/la investigador/a. También hay que tener en cuenta que si aumentamos el número de variables en nuestro modelo, el coeficiente R -cuadrado aumentará, si bien el ajuste puede no haber mejorado en un sentido práctico. Para evitar esta situación, debemos usar los grados de libertad para obtener el estadístico R -cuadrado ajustado.*

CAPÍTULO 5

Ejemplos ilustrativos

En este capítulo presentamos algunos ejemplos para ilustrar cómo se aplica el método que hemos propuesto en el capítulo anterior, aplicados a situaciones reales que han sido ya estudiadas por diversos investigadores en esta materia. De esta forma, podemos comparar los resultados obtenidos empleando nuestro proceso con los obtenidos empleando técnicas anteriores. Antes de ello, repasamos las líneas generales del método propuesto en el capítulo anterior.

5.1. Resumen de la metodología propuesta

Para mayor claridad, antes de describir ejemplos concretos de la metodología que proponemos en cada uno de los casos que pueden ocurrir, resumimos brevemente el procedimiento completo que hemos detallado en el capítulo anterior (recuérdese también la Figura 4.1):

Paso 1. Teniendo en cuenta las características del estudio que se desea realizar, comenzaremos fijando todos los factores que intervienen para poder considerar en el estudio la medida de similitud D y la relación binaria \preceq . Puede ser conveniente que se utilicen aquellas funciones y constantes que hagan que D

sea una auténtica métrica sobre \mathcal{T} y que \preceq sea un auténtico orden parcial sobre \mathcal{T} . En particular, podría ser de mucha utilidad que estos factores llevaran al método de mínimos cuadrados o al método de mínimas desviaciones absolutas.

Paso 2. Generar una muestra aleatoria simple $\{(\underline{X}_i, \mathcal{Y}_i)\}_{i=1}^n$ de la variable conjunta $(\underline{X}, \mathcal{Y})$, y expresar cada uno de los números trapezoidales en función de su centro y de sus amplitudes, es decir, de la forma $\mathcal{Y}_i = \text{Tra}(Y_i, M_i, I_i, S_i)$ para cada $i \in \{1, 2, \dots, n\}$.

Paso 3. Utilizando la medida de similitud D y el software estadístico apropiado para tal tarea, confeccionar tablas de ajustes tanto para la función $Y^c : \text{dom } \underline{X} \rightarrow \mathbb{R}$ que modelizará los centros como para las funciones $M, I, S : \text{dom } \underline{X} \rightarrow \mathbb{R}_0^+$ que describirán las amplitudes (véase, por ejemplo, los resultados de la Tabla 5.1). Dichas tablas de ajustes estarán ordenadas en orden decreciente del coeficiente R^2 (o R^2 -ajustado, según sea el caso). Se tendrán en cuenta dos observaciones muy importantes:

- Si alguno de los procesos de regresión real no produce modelos razonablemente significativos, entonces habrá que ajustar una función constante atendiendo a la metodología desarrollada en la Subsección §4.3.2.
- Entre los muchos modelos significativos que puedan generarse para las funciones que modelizan las amplitudes, habrá que descartar aquellos modelos que puedan tomar valores negativos en el dominio de la variable explicativa $\underline{X} = (X_1, X_2, \dots, X_N)$.

Paso 4. Utilizando cada uno de los posibles emparejamientos de las funciones que pueden modelizar los centros y las amplitudes, construiremos la función estimación

$$\widehat{\mathcal{Z}} = \text{Tra}(\widehat{Y}^c, \widehat{M}, \widehat{I}, \widehat{S}).$$

Paso 5. Para cada una de las posibilidades anteriores, calcularemos el error

difuso total mediante la expresión

$$\mathcal{E} = \sum_{i=1}^n D(\mathcal{Y}_i, \underline{\mathcal{Z}}_{X_i}).$$

Estos resultados se pueden presentar en una tabla (véase, por ejemplo, la Tabla 5.2).

Paso 6. Ordenar los modelos difusos utilizando el orden parcial \preceq introducido en la Definición 3.3.1, teniendo en cuenta los errores difusos obtenidos en el paso anterior para cada uno de los modelos. En los ejemplos, estos resultados se muestran en una tabla denominada *Tabla de ajustes difusos*. Veremos que, aunque pueda parecer lo contrario, el orden establecido entre los modelos en el paso 3 según un coeficiente R -cuadrado real no coincide necesariamente con la ordenación de los modelos obtenida cuando éstos se disponen en este paso de acuerdo con el orden difuso introducido (véase, por ejemplo, la Tabla 5.4).

Obsérvese que los pasos 5 y 6 son la clave para entender que nuestra metodología es claramente diferente de los procedimientos utilizados anteriormente por otros autores: los errores difusos se calculan y se clasifican de acuerdo con un orden parcial difuso sobre el conjunto de todos los números difusos trapezoidales. Por lo tanto, nuestra técnica no se limita al análisis multivariante clásico y, como veremos en los ejemplos, la mejor estimación se consigue con funciones difusas que no se podrían obtener con el simple uso de técnicas reales clásicas.

Una de las propiedades más interesantes de la metodología propuesta es la enorme cantidad de posibilidades que tiene el/la investigador/a para fijar la medida de similitud (o métrica) D y la relación binaria (u orden parcial) \preceq sobre \mathcal{T} . No obstante, en los siguientes ejemplos, por simplicidad y por su sencillez de manejo, supondremos que elegimos los factores necesarios en D y en \preceq para obtener el método de mínimos cuadrados al desarrollar los procesos de regresión reales. De esta forma, salvo que se indique lo contrario, tomaremos $\mathcal{D} = \mathcal{D}_c$, h_1 y h_2 serán las negaciones estándar y tomaremos pseudosemimétricas del estilo $\phi_0(x, y) = |x - y|^{p_0}$, $\psi(x, y) = |x^k - y^k|^p$, $\phi_i(x, y) = |x^{k_i} - y^{k_i}|^{p_i}$ y

$\varphi_i(x, y) = ||x|^{k'_i} - |y|^{k'_i}|^{p'_i}$ cualesquiera que sean x, y en sus respectivos dominios.

5.2. Ejemplo 1: Caso en el que las amplitudes dependen de variables explicativas

En este primer ejemplo del capítulo mostramos la aplicabilidad de la metodología propuesta empleando una situación difusa en la que se pueden observar con detalle cada uno de los pasos descritos anteriormente.

En 2010, Ferraro *et al.* [20] analizaron la relación de dependencia de las ventas del comercio minorista (en millones de dólares) en Estados Unidos durante el año 2002 dependiendo del número de empleados del negocio (véase la Tabla 2 de [20]). La variable independiente X es una variable real que toma valores enteros, ya que se trata del número de empleados/as que trabajan en cada uno de los 22 sectores productivos analizados. Sin embargo, la variable \mathcal{Y} mide las ventas estimadas en cada uno de esos sectores, por lo que no son exactas, y sus valores se reflejan en forma de intervalos. Para desarrollar nuestra metodología, en este estudio identificamos cada uno de esos intervalos con un número difuso triangular cuyo centro coincide con el centro del intervalo dado y cuya amplitud es su radio. Dada la simetría de los números difusos que consideraremos, sólo cabe desarrollar dos procesos de regresión real, uno para los centros y otro para la única amplitud no trivial (en este caso, $M = 0$ e $I = S$). De esta forma, el mismo modelo se aplicará para las amplitudes izquierda y derecha.

Habiendo fijado la métrica D y el orden parcial \preceq , y disponiendo ya de una muestra de la variable conjunta (X, \mathcal{Y}) , desarrollamos procesos de regresión real por mínimos cuadrados para modelizar el centro de los datos y su amplitudes, obteniendo modelos muy significativos para los centros pero no tanto para la amplitud. La Tabla 5.1 muestra un listado, obtenido mediante un paquete estadístico, con los cinco mejores modelos que proporciona el programa tanto

para los centros (con coeficientes R -cuadrado superiores al 90 %) como para las amplitudes (con coeficientes R -cuadrado ligeramente superiores al 60 %). Utilizamos la terminología y la notación que utiliza el propio paquete estadístico para denotar a sus propios modelos.

Tabla 5.1: Tabla de ajustes para el centro (parte superior) y para las amplitudes (parte inferior) izquierda y derecha del ejemplo propuesto en la Sección §5.2.

Modelos propuestos para modelizar la función \widehat{Y}^c		
Modelo	Función	R-cuadrado
Raíz cuadrada de Y	$(29.9912 + 0.000107839 x)^2$	94.80 %
Raíz cuadrada de X	$1807.91 + 1.83701 \cdot 10^{-8} x^2$	94.26 %
Multiplicativo	$0.0155958 x^{0.999421}$	91.58 %
Lineal	$0.0180685 x - 672.731$	91.46 %
Exponencial	$\exp(7.2127 + 0.000002872 x)$	90.62 %
Modelos propuestos para modelizar la función \widehat{S}		
Modelo	Función	R-cuadrado
Multiplicativo	$0.016733 x^{0.878572}$	65.26 %
Raíz cuadrada de Y	$(18.2688 + 0.00004375 x)^2$	64.01 %
Exponencial	$\exp(5.9244 + 0.00000248 x)$	62.47 %
Doble recíproco	$1 / (0.000237192 + 246.094/x)$	61.45 %
Lineal	$0.0034698 x + 185.714$	60.75 %

Al considerar cinco modelos para los centros y otros cinco modelos para la amplitud, encontramos veinticinco posibilidades conjuntas. Para cada uno de los veinticinco posibles emparejamientos, calculamos el error difuso total \mathcal{E} , que queda expresado en la forma $(a/b/c/d)$ (aunque pueda parecer una tarea muy

larga y complicada, un ordenador de gama media tarda menos de 10 segundos en desarrollarla). Los resultados obtenidos se muestran en la Tabla 5.2. Como puede observarse, coherentemente con el hecho de que los números difusos que se emplean son triangulares y simétricos, también los errores difusos totales son triangulares y simétricos. De esta forma, los pasaremos a la notación que resalta su centro y su amplitud, es decir, los expresamos en la forma

$$(a - b/a/a/a + b) = \text{Tra}(a, 0, b, b) = \text{Tri}(a, b).$$

Los resultados de esta conversión pueden observarse en la Tabla 5.3.

Finalmente, la Tabla 5.4 muestra el resultado que se obtiene al ordenar estos modelos conforme al orden parcial difuso establecido. De esta forma, el modelo de regresión difusa que determinamos utilizando el proceso que propusimos en el capítulo anterior es $\widehat{\mathcal{Z}}_x = \text{Tri}(\widehat{Y}_x, \widehat{I}_x, \widehat{S}_x)$, donde

$$\widehat{Y}_x = 1807.91 + 1.83701 \cdot 10^{-8}x^2 \quad \text{e} \quad \widehat{I}_x = \widehat{S}_x = 0.0034698x + 185.714.$$

A continuación comparamos los resultados que hemos obtenido utilizando nuestra metodología con los resultados obtenidos en [20]. Cabe aclarar que la metodología desarrollada en [20], que se nota por MF, emplea una transformación logarítmica de las amplitudes para evitar el problema de asignación de valores negativos a los mismos. Como resultado de su proceso de regresión, se proponen para modelizar el centro y la amplitud las funciones llamadas *lineal* y *Exponencial*, respectivamente, en la Tabla 5.1.

$$\widehat{Y}_x^{\text{MF}} = 0.0180685x - 672.731 \quad \text{e} \quad \widehat{I}_x^{\text{MF}} = \widehat{S}_x^{\text{MF}} = \exp(5.9244 + 0.00000248x).$$

La metodología propuesta en este trabajo, que se nota por MR, indica que los modelos que, conjuntamente, son más adecuados para el centro y las amplitudes son los llamados *cuadrado de X* y *lineal*, respectivamente.

Tabla 5.2: Tabla de errores $\mathcal{E} = \sum_{i=1}^n D(\mathcal{Y}_i, \underline{Z}_{x_i})$, escalados por una factor de 10^{-7} .

Centros	$(29.99 + 0.0001x)^2$	$1807 + 1.84 \cdot 10^{-8}x^2$	$0.0156x^{0.999}$	$0.018x - 672.7$	$e^{2.87 \times 10^{-6}x + 7.21}$
Amplitudes	$(2.74/4.56/6.39)$	$(2.6/4.43/6.25)$	$(6.51/8.34/10.16)$	$(4.76/6.59/8.41)$	$(4.60/6.42/8.24)$
$0.0167x^{0.879}$	$(18.27 + 0.000044x)^2$	$(2.78/4.56/6.35)$	$(2.64/4.43/6.21)$	$(6.55/8.34/10.12)$	$(4.80/6.59/8.37)$
$e^{2.48 \times 10^{-6}x + 5.92}$	$\frac{x}{0.000237x + 246.1}$	$(2.46/4.56/6.67)$	$(2.32/4.43/6.53)$	$(6.23/8.34/10.44)$	$(4.48/6.59/8.69)$
$185.7 + 0.0035x$		$(1.79/4.56/7.34)$	$(1.65/4.43/7.20)$	$(5.56/8.34/11.11)$	$(3.81/6.59/9.36)$
		$(2.88/4.56/6.25)$	$(2.75/4.43/6.11)$	$(6.66/8.34/10.02)$	$(4.91/6.59/8.27)$
				$(4.74/6.42/8.10)$	

Tabla 5.3: Tabla de errores $10^{-7} \mathcal{E} = 10^{-7} \sum_{i=1}^n D(\mathcal{Y}_i, \underline{Z}_{x_i})$.

Centros	$(29.99 + 0.0001x)^2$	$1807 + 1.84 \cdot 10^{-8}x^2$	$0.0156x^{0.999}$	$0.018x - 672.7$	$e^{2.87 \times 10^{-6}x + 7.21}$
Amplitudes	$\text{Tri}(4.56, 1.82)$	$\text{Tri}(4.43, 1.82)$	$\text{Tri}(8.34, 1.82)$	$\text{Tri}(6.59, 1.82)$	$\text{Tri}(6.42, 1.82)$
$0.0167x^{0.879}$	$\text{Tri}(4.56, 1.78)$	$\text{Tri}(4.43, 1.78)$	$\text{Tri}(8.34, 1.78)$	$\text{Tri}(6.59, 1.78)$	$\text{Tri}(6.42, 1.78)$
$(18.27 + 0.000044x)^2$	$\text{Tri}(4.56, 2.11)$	$\text{Tri}(4.43, 2.11)$	$\text{Tri}(8.34, 2.11)$	$\text{Tri}(6.59, 2.11)$	$\text{Tri}(6.42, 2.11)$
$e^{2.48 \times 10^{-6}x + 5.92}$	$\text{Tri}(4.56, 2.78)$	$\text{Tri}(4.43, 2.78)$	$\text{Tri}(8.34, 2.78)$	$\text{Tri}(6.59, 2.78)$	$\text{Tri}(6.42, 2.78)$
$\frac{x}{0.000237x + 246.1}$	$\text{Tri}(4.56, 1.68)$	$\text{Tri}(4.43, 1.68)$	$\text{Tri}(8.34, 1.68)$	$\text{Tri}(6.59, 1.68)$	$\text{Tri}(6.42, 1.68)$
$185.7 + 0.0035x$					

Tabla 5.4: Tabla de ajustes difusos, cuyos modelos han sido ordenados atendiendo al orden parcial \preceq sobre \mathcal{T} .

Centros Amplitudes	$1807 + 1.84 \cdot 10^{-8}x^2$	$(29.99 + 0.0001x)^2$	$e^{2.87 \times 10^{-6}x + 7.21}$	$0.018x - 672.7$	$0.0156x^{0.999}$
$185.7 + 0.0035x$	Tri(4.43, 1.68)	Tri(4.56, 1.68)	Tri(6.42, 1.68)	Tri(6.59, 1.68)	Tri(8.34, 1.68)
	\neq	\neq	\neq	\neq	\neq
$(18.27 + 0.000044x)^2$	Tri(4.43, 1.78)	Tri(4.56, 1.78)	Tri(6.42, 1.78)	Tri(6.59, 1.78)	Tri(8.34, 1.78)
	\neq	\neq	\neq	\neq	\neq
$0.0167x^{0.879}$	Tri(4.43, 1.82)	Tri(4.56, 1.82)	Tri(6.42, 1.82)	Tri(6.59, 1.82)	Tri(8.34, 1.82)
	\neq	\neq	\neq	\neq	\neq
$e^{2.48 \times 10^{-6}x + 5.92}$	Tri(4.43, 2.11)	Tri(4.56, 2.11)	Tri(6.42, 2.11)	Tri(6.59, 2.11)	Tri(8.34, 2.11)
	\neq	\neq	\neq	\neq	\neq
$\frac{x}{0.000237x + 246.1}$	Tri(4.43, 2.78)	Tri(4.56, 2.78)	Tri(6.42, 2.78)	Tri(6.59, 2.78)	Tri(8.34, 2.78)
	\neq	\neq	\neq	\neq	\neq

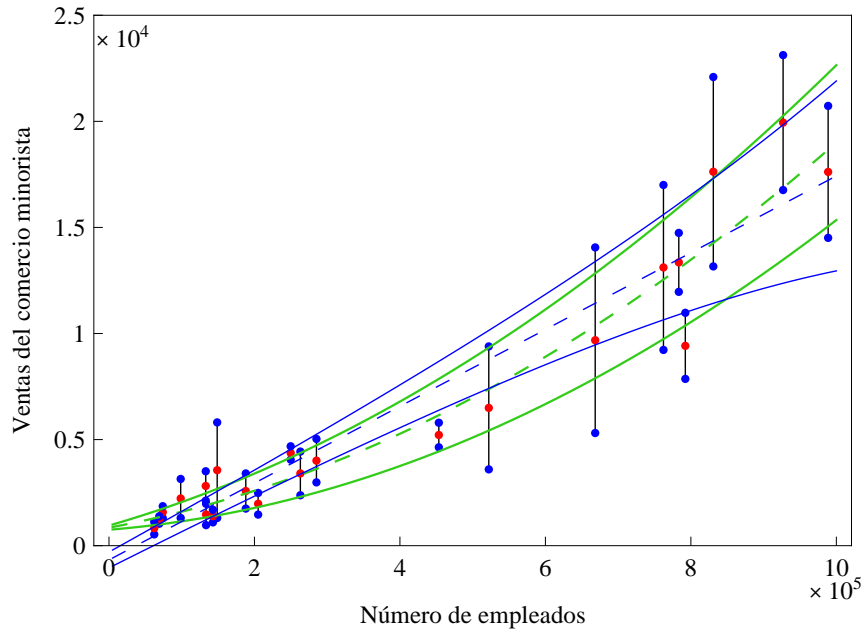
Como se muestra en la Figura 5.1, el modelo MF presenta varios problemas. Cuando el número de empleados es inferior a 6000, las ventas del comercio minorista que predice el modelo MF entran en valores negativos. Observamos también que, debido al comportamiento natural de la función exponencial, el modelo estimado MF tendría graves problemas de modelización cuando el número de empleados aumenta (véase la figura 5.1 b)).

Por otro lado, la Tabla 5.5 incluye los resultados de una comparación de los modelos de regresión difusos MF y MR utilizando SSE y R -cuadrado (recuérdese la Observación 4.4). Nótese que, aunque podríamos haber tenido en cuenta las cuatro esquinas del número difuso rectangular, en este caso, dado que los valores de salida son rectángulos (o números difusos con un centro y una única amplitud significativa), tiene más sentido utilizar una única suma de cuadrados para los centros ($SSE_2 = SSE_3$) y otra para la amplitud (SSE_4 , que sería igual a SSE_1 , que no se ha tenido en cuenta). Finalmente, como conclusión de las medidas de bondad de ajuste calculadas, podemos afirmar que el modelo MR es más apropiado que el modelo MF.

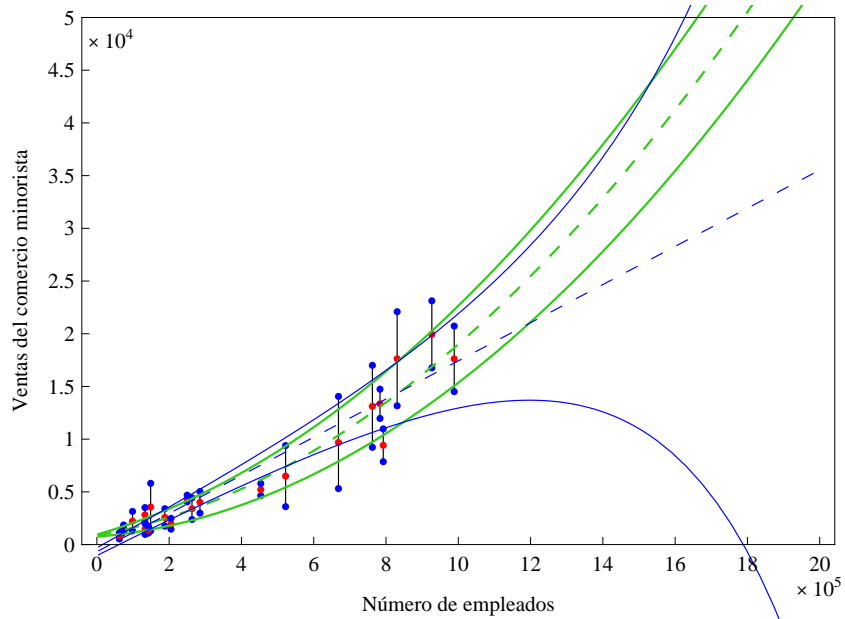
Tabla 5.5: Comparación de SSE y R^2 para los modelos propuestos por MF y MR.

Modelo de regresión difusa	SSE_1	$SSE_2 = SSE_3$	SSE_4
Ferraro et al. [20] (MF)		$7.99227 \cdot 10^7$	$6.58725 \cdot 10^7$
Metodología propuesta (MR)		$3.68159 \cdot 10^7$	$4.42727 \cdot 10^7$
Modelo de regresión difusa	SSE	$\Sigma_j E_j$	\tilde{R}^2
Ferraro et al. [20] (MF)	$9.39716 \cdot 10^7$	$3.05639 \cdot 10^8$	$7.64099 \cdot 10^7$
Metodología propuesta (MR)	$8.53534 \cdot 10^7$	$2.10715 \cdot 10^8$	$5.26787 \cdot 10^7$

Observación 5.2.1 Como acabamos de ver, algunos/as autores/as establecen



(a) Comparación en el dominio de los datos.



(b) Predicción a gran escala.

Figura 5.1: Comparación entre los modelos MF (en color azul) y MR (en color verde).

a priori condiciones de no negatividad sobre las funciones $M_{\underline{x}}$, $I_{\underline{x}}$ y $S_{\underline{x}}$. En este contexto, para el problema de analizar las relaciones de regresión lineal, Coppi et al. [15] propusieron un modelo para datos de entrada reales y respuesta difusa de tipo LR. Los autores consideraron un procedimiento de estimación por mínimos cuadrados iterativo e impusieron una condición de no negatividad en el problema de minimización numérica para evitar amplitudes estimadas negativas. Ferraro et al. [20] propusieron un modelo alternativo para superar la condición de no negatividad, porque las deducciones para los modelos con estas restricciones normalmente hacen el problema más complejo y menos eficiente (véase [20] y las referencias que contiene al respecto). Para superar esta dificultad, emplearon transformaciones logarítmicas sobre los valores de las amplitudes. Nosotros estamos en contra de tales restricciones impuestas desde el principio pues la variedad de modelos que son propuestos actualmente por el software actual permite determinar, casi siempre, modelos significativos que toman valores no negativos en el dominio de la variable de entrada X .

5.3. Ejemplo 2: Caso con datos analizados por muchos investigadores diferentes

En esta sección vamos a introducir un ejemplo que, si bien carece del nivel de detalle y complejidad que se habían introducido en estudios anteriores, tiene la enorme ventaja de que ha sido considerado en la literatura previamente por muchos investigadores (véase [12, 16, 29, 33, 37, 42, 53, 60, 69]) para la evaluación experimental y la comparación de las diferentes metodologías propuestas. Desarrollamos, pues, esta sección, como un estudio comparativo.

Tabla 5.6: Datos de Tanaka *et al.* y sus correspondientes amplitudes.

X	Y^I	Y^c	Y^S	I_i	S_i
1	6.2	8	9.8	1.8	1.8
2	4.2	6.48	6	2.2	2.2
3	6.9	9.5	12.1	2.6	2.6
4	10.9	13.5	16.1	2.6	2.6
5	10.6	13	15.4	2.4	2.4

En 1987, Tanaka *et al.* diseñaron un ejemplo para ilustrar la metodología de regresión difusa que proponían con objeto de tratar el problema en el que hay una única variable exploratoria numérica y una única variable dependiente difusa, la cual tomaba valores difusos triangulares simétricos. En su ejemplo consideraron cinco pares de observaciones que incluimos en la Tabla 5.6.

Tabla 5.7: Tabla de ajustes para el centro (parte superior) y para las amplitudes izquierda y derecha (parte inferior) del ejemplo propuesto en la Sección §5.3.

Modelos propuestos para modelizar la función \widehat{Y}^c		
Modelo	Función	R-cuadrado
Doble cuadrado	$\sqrt{45.4676 + 5.80222x^2}$	78.84 %
Cuadrado de X	$6.97118 + 0.28262 x^2$	77.74 %
Cuadrado de Y	$\sqrt{3.905 + 35.129 x}$	77.27 %
Raíz cuadrada de Y -Cuadrado de X	$(2.65312 + 0.0446295 x^2)^2$	76.42 %
Lineal	$4.95 + 1.71 x$	76.09 %
⋮	⋮	⋮
Modelos propuestos para modelizar la función \widehat{S}		
Modelo	Función	R-cuadrado
Doble recíproco	$1/(0.344065 + 0.208324/x)$	90.24 %
Multiplicativo	$1.87501 \cdot x^{0.212984}$	78.01 %
Raíz cuadrada de $Y - \log X$	$(1.37031 + 0.156125 \log x)^2$	77.04 %
Log X	$1.8806 + 0.458902 \log x$	75.94 %
Lineal	$1.84 + 0.16x$	57.14 %
⋮	⋮	⋮

Tabla 5.8: Tabla de errores $\mathcal{E} = \sum_{i=1}^n D(\mathcal{Y}_i, \underline{Z}_{x_i})$ para el Ejemplo 5.3.

Amplitudes	Centros	$\sqrt{45,47 + 5,8x^2}$	$6,97 + 0,283 x^2$	$\sqrt{3,9 + 35,1 x}$	$(2,65 + 0,045x^2)^2$	$4,95 + 1,71 x$
$1/(0,344 + 0,2/x)$		(8,01/8,09/8,16)	(8,48/8,56/8,63)	(11,15/11,23/11,31)	(9,22/9,29/9,37)	(9,11/9,19/9,26)
$1,87501 \cdot x^{0,212984}$		(7,96/8,09/8,21)	(8,43/8,56/8,68)	(11,11/11,23/11,36)	(9,17/9,29/9,42)	(9,06/9,19/9,31)
$(1,37 + 0,156 \log x)^2$		(7,97/8,09/8,20)	(8,44/8,56/8,67)	(11,12/11,23/11,35)	(9,18/9,29/9,41)	(9,07/9,19/9,30)
$1,88 + 0,46 \log x$		(7,98/8,09/8,19)	(8,45/8,56/8,66)	(11,12/11,23/11,34)	(9,19/9,29/9,40)	(9,08/9,19/9,29)
$1,84 + 0,16x$		(7,89/8,09/8,28)	(8,36/8,56/8,75)	(11,04/11,23/11,42)	(9,10/9,29/9,48)	(9,00/9,19/9,38)

Tabla 5.9: Tabla de errores $\mathcal{E} = \sum_{i=1}^n D(\mathcal{Y}_i, \underline{Z}_{x_i})$ para el Ejemplo 5.3.

Amplitudes	Centros	$\sqrt{45,47 + 5,8x^2}$	$6,97 + 0,283 x^2$	$\sqrt{3,9 + 35,1 x}$	$(2,65 + 0,045x^2)^2$	$4,95 + 1,71 x$
$1/(0,344 + 0,2/x)$		Tri(8,09, 0,08)	Tri(8,56, 0,08)	Tri(11,23, 0,08)	Tri(9,29, 0,08)	Tri(9,19, 0,08)
$1,875 \cdot x^{0,213}$		Tri(8,09, 0,124)	Tri(8,56, 0,124)	Tri(11,23, 0,124)	Tri(9,29, 0,124)	Tri(9,19, 0,124)
$(1,37 + 0,156 \log x)^2$		Tri(8,09, 0,115)	Tri(8,56, 0,115)	Tri(11,23, 0,115)	Tri(9,29, 0,115)	Tri(9,19, 0,115)
$1,88 + 0,46 \log x$		Tri(8,09, 0,108)	Tri(8,56, 0,108)	Tri(11,23, 0,108)	Tri(9,29, 0,108)	Tri(9,19, 0,108)
$1,84 + 0,16x$		Tri(8,09, 0,19)	Tri(8,56, 0,19)	Tri(11,23, 0,19)	Tri(9,29, 0,19)	Tri(9,19, 0,19)

Tabla 5.10: Tabla de ajustes difusos (los modelos han sido ordenados por el criterio difuso) para el Ejemplo 5.3.

Centros Amplitudes	$\sqrt{45,47 + 5,8x^2}$	$6,97 + 0,283 x^2$	$4,95 + 1,71 x$	$(2,65 + 0,045x^2)^2$	$\sqrt{3,9 + 35,1 x}$
$1/(0,344 + 0,2/x)$	Tri(8,09, 0,08)	Tri(8,56, 0,08)	Tri(9,19, 0,08)	Tri(9,29, 0,08)	Tri(11,23, 0,08)
	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle
$1,88 + 0,46 \log x$	Tri(8,09, 0,108)	Tri(8,56, 0,108)	Tri(9,19, 0,108)	Tri(9,29, 0,108)	Tri(11,23, 0,108)
	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle
$(1,37 + 0,156 \log x)^2$	Tri(8,09, 0,115)	Tri(8,56, 0,115)	Tri(9,19, 0,115)	Tri(9,29, 0,115)	Tri(11,23, 0,115)
	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle
$1,875 \cdot x^{0,213}$	Tri(8,09, 0,124)	Tri(8,56, 0,124)	Tri(9,19, 0,124)	Tri(9,29, 0,124)	Tri(11,23, 0,124)
	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle	\sphericalangle
$1,84 + 0,16x$	Tri(8,09, 0,19)	Tri(8,56, 0,19)	Tri(9,19, 0,19)	Tri(9,29, 0,19)	Tri(11,23, 0,19)

Tabla 5.11: Tabla de modelos difusos (ordenados por el criterio difuso) para el Ejemplo 5.3.

Autores	Modelo de regresión difusa	SSE	$\Sigma_j E_j$	\tilde{R}^2
Tanaka <i>et al.</i> (1987)	$\hat{\mathcal{Y}}_{THW}(x) = (0, 3.85, 7.7) + 2.1x$	67.235	16.8088	56.51 %
Diamond (1988)	$\hat{\mathcal{Y}}_{DM}(x) = (3.11, 4.95, 6.79) + (1.55, 1.71, 1.87)x$	37.132	9.283	75.98 %
Kim y Bishu (1998)	$\hat{\mathcal{Y}}_{KB}(x) = (3.11, 4.95, 6.84) + (1.55, 1.71, 1.82)x$	37.207	9.30175	75.93 %
Wu y Tseng (2002)	$\hat{\mathcal{Y}}_{WT}(x) = (3.11, 4.95, 6.79) + (1.55, 1.71, 1.87)x$	37.132	9.283	75.98 %
Kao y Chyu (2003)	$\hat{\mathcal{Y}}_{KC}(x) = 4.81 + 1.72x + (-2.2, 0.12, 2.44)$	37.65	9.4125	75.65 %
Nasrabadi y Nasrabadi (2004)	$\hat{\mathcal{Y}}_{NN}(x) = (2.36, 4.86, 7) + 1.73x$	38.542	9.6355	75.07 %
Hojati <i>et al.</i> (2005)	$\hat{\mathcal{Y}}_{HBS}(x) = (5.1, 6.75, 8.4) + (1.1, 1.25, 1.4)x$	49.61	12.4025	67.91 %
Chen y Hsueh (2009)	$\hat{\mathcal{Y}}_{CH}(x) = 1.71x + (2.63, 4.95, 7.27)$	37.644	9.411	75.65 %
Roldán <i>et al.</i> (2012)	$\hat{\mathcal{Y}}_{RRM}(x) = \text{Tri} \left(4.95 + 1.71x, \frac{1}{0.344+0.208/x} \right)$	36.9017	9.22542	76.13 %
Nuestra aproximación	$\hat{\mathcal{Y}}_{RRMA}(x) = \text{Tri} \left(\sqrt{45.47 + 5.8x^2}, \frac{1}{0.344+0.208/x} \right)$	32.4945	8.1262	78.98 %

La Tabla 5.7 nos da un listado de algunos de los mejores modelos para centros y amplitudes, ordenada utilizando el criterio de R -cuadrado. La Tabla 5.8 presenta los errores difusos totales \mathcal{E} , calculados de la forma $(a/b/c)$. Para la correcta ordenación de estos errores difusos por el orden parcial introducido, se muestra en la Tabla 5.9 sus correspondientes expresiones en la forma $\text{Tri}(A^c, A^s)$ (teniendo en cuenta que son simétricos). Entonces, en la Tabla 5.10, se presentan todos los modelos ordenados atendiendo al criterio difuso propuesto. Es muy importante tener en cuenta que este ranking no coincide con el criterio de R -cuadrado.

Finalmente, el modelo propuesto por nuestra metodología utilizando el ordenamiento difuso anterior es

$$\hat{Z}_x = \text{Tri} \left(\sqrt{45,4676 + 5,80222x^2}, \frac{1}{0,344065 + 0,208324/x} \right)$$

para cada $x \in \text{dom } X$.

La Tabla 5.11 muestra los resultados del estudio comparativo entre nuestro modelo y los modelos previos obtenidos por otros autores, utilizando SSE , $\Sigma_j E_j$ y \tilde{R}^2 .

Como conclusión, podemos decir que el enfoque puramente difuso propuesto en esta Memoria mejora significativamente los resultados obtenidos por otros autores.

5.4. Ejemplo 3: Caso en el que las amplitudes no dependen de las variables explicativas

Las instituciones académicas utilizan los datos de satisfacción de los/as estudiantes para comprender, cambiar y mejorar el entorno del campus, creando así ambientes más propicios para el desarrollo de los estudiantes. En este sentido, la satisfacción del estudiante es un indicador de la capacidad de respuesta

de la institución a las necesidades de los estudiantes y una medida de la eficacia, el éxito y la vitalidad institucionales. La mayoría de los cuestionarios que recogen opiniones y juicios de los/as estudiantes se diseñan utilizando una escala Likert de cinco puntos o una escala de números enteros. Sin embargo, como las evaluaciones de opinión/valoración son datos intrínsecamente imprecisos, varios estudios sugieren que podría ser más razonable y factible el uso de una escala de números difusos para evitar la pérdida de información y, por lo tanto, es más conveniente utilizar métodos de regresión difusos en lugar de los métodos de regresión tradicionales.

El objetivo principal de esta investigación era el de conocer si existe alguna relación entre la nota obtenida por cada alumno/a en la prueba inicial y el grado de satisfacción del alumnado al terminar sus estudios en la asignatura de Estadística. En concreto, se trataba de comprobar si la nota inicial puede explicar, en mayor o menor medida, el grado de satisfacción con respecto a la asignatura. De esta forma, la variable explicativa X , que era la nota obtenida por el alumnado en la prueba inicial, es una variable real continua que toma valores en el intervalo real $[0, 10]$. Por otro lado, la variable explicada \mathcal{Y} , que medía el grado de satisfacción del alumnado al terminar sus estudios en la asignatura de Estadística, es una variable aleatoria que toma valores difusos trapezoidales con soporte contenido en el intervalo $[0, 100]$ (siendo 0 la mínima satisfacción y 100, la máxima).

La investigación se llevó a cabo en el segundo cuatrimestre del curso 2011/2012, en la Escuela Técnica Superior de Ciencias de la Información y Telecomunicaciones de la Universidad de Granada, entre el alumnado de Estadística, asignatura del primer curso de la titulación de Ingeniería Informática. Al principio del curso, el profesor explicó el propósito del estudio y dio instrucciones para valorar, de manera anónima, la satisfacción del alumnado respecto de esta asignatura (lo cual se haría al final del cuatrimestre). La prueba inicial

se desarrolló en el mes de febrero de 2012, y se valoró con un número real en el intervalo $[0, 10]$. Para garantizar el anonimato, durante la misma se entregó a cada estudiante un número personalizado. Al terminar sus estudios de dicha asignatura, durante el mes de junio de 2012, se pidió a los estudiantes que respondiesen a varias preguntas, utilizando para ello números difusos con soporte compacto contenido en el intervalo $[0, 100]$. En dicho cuestionario, el alumnado debía reflejar su número personalizado, que sería independiente del examen final de la asignatura. Se dio una semana para recoger las respuestas de satisfacción de los estudiantes. El número de estudiantes matriculados en la asignatura que participaron en la investigación fue de 54. Por sexos, un 11.11 % eran mujeres, mientras que un 88.89 % eran varones. Los datos obtenidos en este estudio se muestran en la Tabla 5.12.

Aplicando el análisis de regresión clásico por mínimos cuadrados a través de un paquete estadístico, podemos encontrar una lista de posibles funciones para modelizar la función Y^c que determinará los centros, resultando éstas ser significativas. Sin embargo, en este caso, ningún modelo puede ser considerado significativo para los datos de las amplitudes y , en consecuencia, el método de mínimos cuadrados no parece el más razonable. La Tabla 5.13 resume los mejores modelos que proporciona el programa estadístico ordenados según su coeficiente R -cuadrado.

Tabla 5.12: Tabla de datos correspondiente al ejemplo de la Sección §5.4.

X_i	Y_i^1	Y_i^2	Y_i^3	Y_i^4	X_i	Y_i^1	Y_i^2	Y_i^3	Y_i^4
8.2	60	70	75	82	2.2	37	43	50	60
2.1	33	47	50	58	8	85	90	90	95
1.9	30	35	40	45	2.8	40	50	53	63
5	57	63	65	75	6	65	75	78	85
5	57	68	73	85	3.1	43	50	55	62
3.1	45	48	50	60	5	55	63	70	82
3.8	50	55	58	70	2.2	38	50	54	63
5	60	68	70	81	9	88	94	96	100
8.5	75	87	89	97	6.6	70	74	75	87
6.5	75	81	83	90	6	70	74	75	78
3.3	46	50	53	55	3.5	50	60	62	75
7.5	72	77	80	85	5.3	45	60	63	71
1.5	32	45	55	58	7.2	70	77	83	90
7	75	84	87	100	1.2	35	40	45	53
2.5	50	60	64	75	1.3	27	30	40	45
6	70	75	75	80	5	60	65	70	75
2.6	44	50	55	58	3.8	50	55	60	65
1.5	32	44	47	55	9.5	80	90	95	100
3.3	45	60	60	70	5.5	60	73	75	85
5	60	67	70	75	6	63	65	72	80
6	65	70	72	80	8	87	90	92	96
8	73	85	88	98	1.2	20	27	30	35
1.4	30	33	35	45	5.1	57	64	67	79
2.5	40	46	48	55	2.7	44	55	58	70
7	70	75	77	80	5.5	60	65	68	77
6.2	70	80	85	92	5	55	63	65	75
9.3	80	90	95	100					

Tabla 5.13: Tabla de ajustes para el centro (parte superior) y para las amplitudes (parte inferior) del ejemplo propuesto en la Sección §5.4.

Modelos propuestos para modelizar la función \widehat{Y}^c		
Modelo	Función	R-cuadrado
Cuadrado de Y	$\sqrt{862.2x + 329.6}$	91.30 %
Raíz cuadrada de X	$27.68\sqrt{x} + 6.22$	91.19 %
Lineal	$6.64x + 32.93$	90.73 %
Raíz cuadrada doble	$(1.77\sqrt{x} + 4.24)^2$	90.66 %
Multiplicativo	$33.23x^{0.446}$	89.57 %
Mejores modelos obtenidos para las amplitudes \widehat{M} , \widehat{I} y \widehat{S}		
Diferencial	Función	R-cuadrado
Para \widehat{M}	$\sqrt{1.218 + 10.197/x}$	14.98 %
Para \widehat{I}	$1/(0.154 + 0,0322/x)$	0.55 %
Para \widehat{S}	$\sqrt{80.905 - 0.4197x^2}$	4.70 %

Tabla 5.14: Tabla de errores considerando diferentes medidas de error difuso para estimar las amplitudes.

	$p = 2, k = 1$	$p = 1, k = 3$	$p = 0.5, k = 2.5$
$\widehat{D}_c^* \mathcal{Z}$	Tra(0, 1.7453, 7.6415, 7.7358)	Tra(0, 1.5, 7, 8)	Tra(0, 1.5, 5, 7)
$\sqrt{862.2x + 329.6}$	Tra(1251.4, 59.3, 604.2, 472.3)	Tra($2.64 \cdot 10^6$, 579.3, 31434, 25910)	Tra(3454.8, 91.15, 552, 554.5)
$27.68\sqrt{x} + 6.22$	Tra(1258.1, 59.3, 604.2, 472.3)	Tra($2.72 \cdot 10^6$, 579.3, 31434, 25910)	Tra(3524.8, 91.15, 552, 554.5)
$6.64x + 32.93$	Tra(1323.4, 59.3, 604.2, 472.3)	Tra($2.65 \cdot 10^6$, 579.3, 31434, 25910)	Tra(3454.6, 91.15, 552, 554.5)
$(1.77\sqrt{x} + 4.24)^2$	Tra(1234.9, 59.3, 604.2, 472.3)	Tra($2.57 \cdot 10^6$, 579.3, 31434, 25910)	Tra(3409.1, 91.15, 552, 554.5)
$33.23x^{0.446}$	Tra(1280.1, 59.3, 604.2, 472.3)	Tra($2.81 \cdot 10^6$, 579.3, 31434, 25910)	Tra(3570.7, 91.15, 552, 554.5)
\vdots	\vdots	\vdots	\vdots

Dado que las amplitudes no dependen de la variable explicativa, nuestra metodología propone ajustar una función constante para estimarlos, conforme a lo explicado en la Subsección §4.3.2. En este caso, nos planteamos el uso de una pseudosemimétrica del tipo

$$\phi(x, y) = |x^k - y^k|^p,$$

donde $k, p > 0$ son variables fijadas por la persona encargada de la investigación. Para estudiar la influencia de dichas constantes, presentamos, a continuación, tres casos posibles. La Tabla 5.14 muestra los errores difusos totales obtenidos al elegir diferentes posibilidades para la función que modeliza los centros (primera columna) y para las constantes p y k : en la segunda columna, $p = 2$ y $k = 1$ (método de mínimos cuadrados), en la tercera, $p = 1$ y $k = 3$, y en la cuarta, $p = 0.5$ y $k = 2.5$. En la segunda fila de dicha tabla escribimos el modelo que resulta más adecuado para $\widehat{\mathcal{D}}_c^* \widehat{\mathcal{Z}}$ (es decir, para \widehat{M} , \widehat{I} y \widehat{S}) dependiendo de los diferentes valores de p y de k .

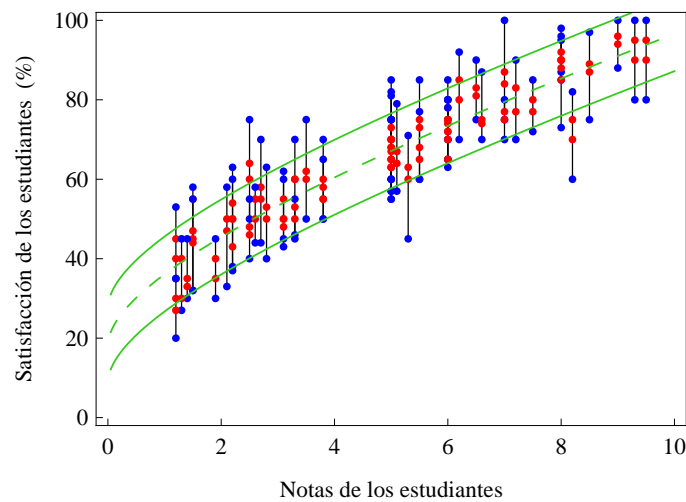


Figura 5.2: Datos de satisfacción observados según la calificación del estudiante y el modelo de regresión difusa estimado

$$\widehat{\mathcal{Z}}_x = \text{Tra}((1.76654\sqrt{x} + 4.24198)^2, 1.745, 7.642, 7.736).$$

Como conclusión, observamos que, en todos los casos estudiados (donde cambian p y k), la mejor opción para modelizar el centro no es la función con mayor coeficiente R -cuadrado real. En este caso, es mejor el modelo

$$\widehat{Y}^c = \widehat{\mathcal{D}}_c \widehat{\mathcal{Z}}_x = (1.76654\sqrt{x} + 4.24198)^2.$$

La Figura 5.2 traza el modelo de regresión difusa obtenida para el caso $p = 2$ y $k = 1$, que viene dado por $\widehat{\mathcal{Z}}_x = \text{Tra}(\widehat{Y}_x, \widehat{M}, \widehat{I}, \widehat{S})$, donde

$$\begin{aligned} \widehat{Y}_x &= (1.76654\sqrt{x} + 4.24198)^2, \\ \widehat{M} &= 1.745283, \quad \widehat{I} = 7.641509 \quad \text{y} \quad \widehat{S} = 7.735849. \end{aligned}$$

5.5. Ejemplo 4: Caso en el que las variables de entrada son difusas

En [67], Wu propuso un enfoque para la construcción de un modelo de regresión difusa. Para ilustrar su metodología, utilizó un ejemplo en el que dos variables de entrada difusas \mathcal{X}_1 y \mathcal{X}_2 trataban de explicar el comportamiento de una variable de respuesta difusa \mathcal{Y} , tomando todas ellas valores difusos triangulares. Los datos que emplearon están recogidos en la Tabla 5.15. Más tarde, Chen y Hsueh [12] volvieron a considerar los mismos datos para explicar su procedimiento de regresión y compararon sus resultados con los resultados obtenidos anteriormente por el enfoque de Wu (véase la Tabla III en [12]).

En este ejemplo, aplicamos nuestra metodología para explicar la relación de dependencia entre la variable aleatoria triangular difusa \mathcal{Y} y las dos variables aleatorias difusas triangulares \mathcal{X}_1 y \mathcal{X}_2 .

De acuerdo con lo que proponemos en la Subsección §4.3.3, este problema debe afrontarse expresando la variable difusa \mathcal{Y} en función de las seis variables

Tabla 5.15: Datos originales utilizados por Wu [67] y por Chen y Hsueh [12], que se analizan en la Sección §5.5.

Variable explicativa 1 $X_1 = (X_1^\ell/X_1^c/X_1^s)$	Variable explicativa 2 $X_2 = (X_2^\ell/X_2^c/X_2^s)$	Variable respuesta Y
(151/274/322)	(1432/2450/3461)	(111/162/194)
(101/180/291)	(2448/3254/4463)	(88/120/161)
(221/375/539)	(2592/3802/5116)	(161/223/288)
(128/205/313)	(1414/2838/3252)	(83/131/194)
(62/86/112)	(1024/2347/3766)	(51/67/83)
(132/265/362)	(2136/3782/5091)	(124/169/213)
(66/98/152)	(1687/3008/4325)	(62/81/102)
(151/330/463)	(1524/2450/3864)	(138/192/241)
(115/195/291)	(1216/2137/3161)	(82/116/159)
(35/53/71)	(1432/2560/3782)	(41/55/71)
(307/430/584)	(2592/4020/5562)	(168/252/367)
(284/372/498)	(2792/4427/6163)	(178/232/346)
(121/236/370)	(1734/2660/4094)	(111/144/198)
(103/157/211)	(1426/2088/3312)	(78/103/148)
(216/370/516)	(1785/2605/4042)	(167/212/267)

aleatorias reales X_1^ℓ , X_1^c , X_1^s , X_2^ℓ , X_2^c y X_2^s que componen las variables difusas $\mathcal{X}_1 = (X_1^\ell/X_1^c/X_1^s)$ y $\mathcal{X}_2 = (X_2^\ell/X_2^c/X_2^s)$ (observemos que, en este caso, $X_1^m = X_2^m = 0$).

Siguiendo la metodología que hemos explicado anteriormente, el modelo de

Tabla 5.16: Comparación de los errores de estimación de varios modelos usando SSE (sección §5.5).

Autores	Modelo de regresión difusa ajustado	<i>SSE</i>
Wu	$3.453 + 0.496\mathcal{X}_1 + 0.009\mathcal{X}_2$	3674.819
Chen-Hsueh	$0,507\mathcal{X}_1 + 0.009\mathcal{X}_2 - (-18.167/0.060/10.592)$	2217.535
Nuestra metodología	$\hat{\mathcal{Z}}_{\underline{X}} = \text{Tri}(\hat{Y}_{\underline{X}}^c, \hat{I}_{\underline{X}}, \hat{S}_{\underline{X}})$ (véase (5.1)-(5.3))	1784.483

regresión difusa ajustado es $\hat{\mathcal{Z}}_{\underline{X}} = \text{Tri}(\hat{Y}_{\underline{X}}^c, \hat{I}_{\underline{X}}, \hat{S}_{\underline{X}})$, donde

$$\begin{aligned} \hat{Y}_{\underline{X}}^c &= 2.5246 + 0.0068X_1^\ell + 0.5269X_1^c - 0.0248X_1^s \\ &\quad + 0.0006X_2^\ell + 0.0043X_2^c + 0.0034X_2^s, \end{aligned} \quad (5.1)$$

$$\begin{aligned} \hat{I}_{\underline{X}} &= 15.3986 + 0.0411X_1^\ell + 0.1892X_1^c - 0.03489X_1^s \\ &\quad - 0.0041X_2^\ell + 0.021X_2^c - 0.016X_2^s, \end{aligned} \quad (5.2)$$

$$\begin{aligned} \hat{S}_{\underline{X}} &= 4.5245 + 0.5813X_1^\ell - 0.4936X_1^c + 0.2551X_1^s \\ &\quad - 0.0092X_2^\ell + 0.0154X_2^c - 0.0080X_2^s, \end{aligned} \quad (5.3)$$

cuyos valores de R -cuadrados ajustados son 99.8234 %, 91.9099 % y 88.9876 %, respectivamente. La Tabla 5.16 muestra que el error de estimación total SSE de nuestra propuesta es el más bajo: 1784.483. Por último, obsérvese que el valor R -cuadrado del modelo que hemos obtenido con la metodología propuesta es de $\tilde{R}^2 = 0.9909$.

5.6. Ejemplo 5: Caso en el que el modelo propuesto explica el 100% de la variabilidad difusa observada

En algunos casos, nuestra metodología puede producir resultados excelentes. Por ejemplo, consideremos el caso experimental siguiente, que fue presentado por Kim y Bishu en [37], y que más tarde ha sido ampliamente estudiado, entre otros, en [12, 14, 28, 41, 58]. En este ejemplo, tenemos en cuenta los tiempos de respuesta cognitivos del personal de sala de la planta de energía nuclear en un suceso anormal (más detalles se pueden encontrar en [37]). La Tabla 5.17 presenta los datos originales. Un modelo difuso (por lo general, lineal) se asume entre el tiempo de respuesta cognitiva y la experiencia de las personal dentro de la sala de control (en años) y la experiencia fuera de una sala de control (en años) y la educación (en años).

Las observaciones de las variables independientes son datos reales ($\underline{X} = (X_1, X_2, X_3)$) aunque las observaciones de la variable dependiente (\mathcal{Y}) se presentan como números difusos triangulares simétricos (por tanto, $M = 0$ e $I = S$). Aplicando la metodología que proponemos, se obtiene el modelo de regresión difusa $\hat{\underline{Z}}_{\underline{X}} = \text{Tri}(\hat{Y}_{\underline{X}}^c, \hat{S}_{\underline{X}})$, donde

$$\begin{aligned}\hat{Y}_{\underline{X}}^c &= \exp(-788.318 - 9.20165 X_1 - 3.9764 X_2 + 111.28 X_3 \\ &\quad + 0.889021 X_1^2 + 0.746288)X_2^2 - 3.92647X_3^2 + 14.2274\sqrt{X_1}); \\ \hat{S}_{\underline{X}} &= \exp(-791.587 - 9.18783 X_1 - 3.98627 X_2 + 111.799 X_3 \\ &\quad + 0.887756 X_1^2 + 0.748419)X_2^2 - 3.94458X_3^2 + 14.2211\sqrt{X_1}).\end{aligned}$$

Este modelo no sólo es bueno, sino que es el mejor posible, ya que el conjunto de datos se ajusta perfectamente a este modelo ($\tilde{R}^2 = 100\%$). Por lo tanto, no es necesario compararlo con los modelos obtenidos en [12, 14, 28, 37, 41, 58].

Tabla 5.17: Datos introducidos por Kim y Bishu en [37].

	Años de experiencia		Años de experiencia	Educación	Tiempo de
	dentro de control (X_1)	fuera de control (X_2)	(X_3)	respuesta (Y)	
Equipo 1	2.0	0.0	15.25	Tri(5.83, 3.56)	
Equipo 2	0.0	5.0	14.13	Tri(0.85, 0.52)	
Equipo 3	1.13	1.5	14.13	Tri(13.93, 8.5)	
Equipo 4	2.0	1.25	13.63	Tri(4, 2.44)	
Equipo 5	2.19	3.75	14.75	Tri(1.65, 1.01)	
Equipo 6	0.25	3.5	13.75	Tri(1.58, 0.96)	
Equipo 7	0.75	5.25	15.25	Tri(8.18, 4.99)	
Equipo 8	4.25	2.0	13.5	Tri(1.85, 1.13)	

CAPÍTULO 6

Conclusiones y observaciones finales

Muchos trabajos han propuesto diferentes nociones de medida de la distancia entre los números difusos, que han demostrado su aplicabilidad en estudios empíricos. Sin embargo, para dar un número real con el fin de describir la similitud entre dos números difusos implica una gran pérdida de información importante. Además, en general, las medidas que se proponen, en general, no verifican los cuatro axiomas clásicos de una métrica. Por otro lado, tan sólo unos pocos de trabajos interpretaban la medida de distancia entre dos números difusos como otro número difuso.

Para superar estos inconvenientes, esta memoria presenta una nueva familia de medidas de distancia entre números difusos que da como resultado otro número difuso. Para resolver a este problema, previamente, hemos introducido una noción de métrica sobre un conjunto arbitrario que toma valores en el mismo conjunto que es una relación binaria reflexiva y transitiva, entre los números difusos que, en el caso particular de que los números difusos que estamos considerando sean trapezoidales, es un orden parcial. Este orden parcial nos permite considerar números difusos no negativos en la que nuestra familia de medidas de distancia toma valores.

Las principales ventajas de la familia de medidas de similitud introducida son los siguientes:

1. se puede medir la similitud entre dos números difusos arbitrarios y no se restringe sólo a números difusos triangulares o trapezoidales;
2. la medida de la distancia entre dos números difusos arbitrarios es también un número difuso;
3. esta familia depende de una amplia gama de diferentes valores (incluyendo constantes, defuzzificaciones y pseudosemimétricas) que permite al científico tener distintas posibilidades de medición;
4. el investigador puede controlar desde el principio el tipo de número difuso que se puede obtener;
5. esta metodología extiende el caso real de tal manera que podemos considerar, en el mismo estudio, números difusos y reales;
6. si el investigador considera que es importante, es posible manejar una única clase de números difusos (por ejemplo, el números difusos triangulares), ya que, bajo ciertas condiciones generales, la medida de distancia D se puede aplicar a ciertas clases de números difusos en sí mismas;
7. una vez que hemos fijado los elementos geométricos que utilizaremos en la medida de distancia D (es decir, q_1 , q_2 , \mathcal{D} , h_1 , h_2 , ϕ_1 , ϕ_3 , ϕ_4 y ψ_2 dependiendo del interés del investigador), el cálculo de $D(A, B)$ es un proceso muy fácil, intuitivo y con un coste computacional muy bajo.

La metodología descrita se ilustra para comparar nuestros resultados con los obtenidos por otros procedimientos anteriores más complejos, demostrando así el potencial de posibles aplicaciones a la clasificación de números difusos.

Utilizando esta familia, en esta memoria ha explorado el problema de analizar relaciones de regresión difusa lineales y no lineales entre una variable respuesta difusa y variables explicativas que son caracteres cuantitativos. La mayoría de los investigadores usan formas simples de números difusos en estudios de la vida real porque son más fáciles de fijar y manejar. Por lo tanto, en esta memoria la imprecisión de la variable respuesta está descrita por números difusos trapezoidales.

En general, cuando las variables aleatorias difusas se consideran en análisis de regresión para construir las relaciones entre las variables respuesta y explicativas, este problema se vuelve más complejo y podemos encontrar dificultades en encontrar soluciones óptimas, especialmente para problemas no lineales. Como la formulación propuesta un método ampliamente utilizado y conocido, el método de mínimos cuadrados, la principal contribución de esta formulación es ser capaz de obtener los parámetros difusos fácilmente, dando resultados que son compatibles con aquellos obtenidos por técnicas más complicadas tal y como se muestra en los ejemplos.

Observar que no es necesario imponer restricciones de no negatividad y que si no encontramos un modelo que se ajuste de forma significativa a los datos, en algunas situaciones es válido utilizar una constante para la estimación del mismo, como hemos visto en los ejemplos.

Finalmente el uso de nuevas distancias difusas nos permite ser consistentes con el tipo de datos considerados.

Resumiendo, las principales ventajas del método de regresión difusa que presentamos son:

- proporciona a los investigadores un enfoque relativamente sencillo para el problema de analizar las relaciones de regresión cuando los datos observados pueden verse afectados por diferentes fuentes de incertidumbre;

- en nuestro estudio, no es necesario tener en cuenta restricciones de no negatividad desde el principio;
- los investigadores pueden estudiar relaciones estadísticas entre variables difusas con un método que se puede aplicar en diferentes contextos;
- la técnica no se limita a considerar modelos lineales;
- no se limita a intervalos numéricos o variables fuzzy triangulares (como en otras investigaciones);
- la metodología no se limita a variables explicativas reales;

Pensamos que es necesario realizar investigaciones futuras para encontrar nuevas propiedades de la familia de medidas de similitud introducidas y cómo aplicar esta familia a contextos reales. En trabajos futuros, estamos interesados en la descripción de nuevas familias más generales de números difusos en las que algunos casos particulares de D sean distancias en el sentido de la definición 3.1.2, y también en la sustitución del intervalo $\mathbb{I} = [0, 1]$ por un retículo más general L (véase [49, 50]). Además, estamos interesados en la aplicación de nuestras medidas de distancia en la resolución de problemas estadísticos en los que esté presente la incertidumbre y sea necesario el uso de una medida de distancia para su solución.

Bibliografía

- [1] M. Adabitar Firozja, G.H. Fath-Tabar, Z. Eslampia, The similarity measure of generalized fuzzy numbers based in interval distance, *Applied Mathematics Letters* 25 (2012) 1528–1534.
- [2] C. Aguilar-Peña, A.F. Roldán-López-de-Hierro, C. Roldán-López-de-Hierro, J. Martínez-Moreno, Estimation of a fuzzy regression model using fuzzy distances, aceptado en mayo de 2015 para su publicación en *IEEE Transactions on Fuzzy Systems*.
- [3] T. Allahviranloo, M.A. Firozja, Ranking of fuzzy numbers by a new metric, *Soft Computing* 14 (2010) 773–783.
- [4] T. Allahviranloo, R. Nuraei, M. Ghanbari, E. Haggi, A.A. Hossein-zadeh, A new metric for L - R fuzzy numbers and its application in fuzzy linear systems, *Soft Computing* 16 (2012) 1743–1754.
- [5] A.I. Ban, L. Coroianu, Nearest interval, triangular and trapezoidal approximation of a fuzzy number preserving ambiguity, *International Journal of Approximate Reasoning* 53 (2012) 805–836.
- [6] J.J. Buckley, L.J. Jowers, Monte Carlo methods in fuzzy optimization, *Studies in Fuzziness and Soft Computing*, Volume 222, Springer, 2008.

- [7] F. Castro-Company, S. Romaguera, Experimental results for information system based on accesses locality via intuicionist fuzzy metrics, *Open Cybern. Syst. J.* 2 (2008) 158–172.
- [8] C.T. Chang, An approximation approach for representing S-shaped membership functions, *IEEE Transactions on Fuzzy Systems*, 18 (2) (2010), art. no. 5411779, 412–424.
- [9] S.H. Chatterjee, S. Ali, B. Price, *Regression Analysis by Example*. 5th Edition. Wiley (2012).
- [10] S.M. Chen, New methods for subjective mental workload assessment and fuzzy risk analysis, *Cybernetics and Systems* 27 (1996) 449–472.
- [11] S.J. Chen, S.M. Chen, Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers, *IEEE Transactions on Fuzzy Systems* 11 (2003) 45–56.
- [12] L.H. Chen, C.C. Hsueh, Fuzzy regression models using the least-squares method based on the concept of distance, *IEEE Transactions on Fuzzy Syst* 17 (2009) 1259–1272.
- [13] C.H. Cheng, A new approach for ranking fuzzy numbers by distance method, *Fuzzy Sets and Systems* 95 (1988) 307–317.
- [14] S.H. Choi, J. J. Buckley, Fuzzy regression using least absolute deviation estimators, *Soft Comput.* 12 (2008) 257–263.
- [15] R. Coppi, P. D’Urso, P. Giordani, and A. Santoro, Least squares estimation of a linear regression model with LR fuzzy response, *Comput. Stat. Data Anal.* 51 (2006) 267–286.
- [16] P. Diamond, Fuzzy least squares. *Inform Sci* 46 (1988) 141–157.

-
- [17] D. Dubois, H. Prade, Operations on fuzzy numbers, *International Journal of System Sciences* 9 (1978) 613–626.
- [18] D. Dubois, H. Prade, *Possibility Theory*, New York, Plenum Press, 1988.
- [19] D. Dubois, H. Prade, Fuzzy elements in a fuzzy set. *Proc. 10th Internat. Fuzzy Systems Assoc. (IFSA) Congr. 2005*, 55-60, Beijing, Springer.
- [20] M.B. Ferraro, R. Coppi, G. González-Rodríguez, and A. Colubi, A linear regression model for imprecise response, *Int. J. Approx. Reasoning* 51 (2010) 759–770.
- [21] F. Galton, Regression towards mediocrity in hereditary stature, *Journal of the Anthropological Institute* 15 (1886) 246-263
- [22] T. Ganesan, P. Vasant, I. Elamvazuthi, Hybrid PSO approach for solving non-convex optimization problems, *Archives of Control Sciences* 22 (1) (2012) 5–23.
- [23] A. George, P. Veeramani, On some results in fuzzy metric spaces, *Fuzzy Sets Syst.* 64 (1994) 395–399.
- [24] R. Goetschel, W. Voxman, Elementary fuzzy calculus, *Fuzzy Sets and Systems* 18 (1986) 31–43.
- [25] S. Gorard, Revisiting a 90-year-old debate: the advantages of the mean deviation, *British Educational Research Association Annual Conference*, University of Manchester (2004) 16–18.
- [26] P. Grzegorzewski and E. Mrówka, Trapezoidal approximations of fuzzy numbers, *Fuzzy Sets Syst.* 153 (2005) 115–135.
- [27] D. Guha, D. Chakraborty, A new approach to fuzzy distance measure and similarity measure between two generalized fuzzy numbers, *Applied Soft Computing* 10 (2010) 90–99.

- [28] H. Hassanpour, H.R. Maleki, M.A. Yaghoobi, A goal programming approach to fuzzy linear regression with non-fuzzy input and fuzzy output data, *Asia-Pacific Journal of Operational Research* 26 (2009) 587–604.
- [29] M. Hojati, C.R. Bector, K. Smimou, A simple method for computation of fuzzy linear regression, *Eur J Oper Res*, 166 (2005) 172–184.
- [30] D.H. Hong, C. Hwang, Support vector fuzzy regression machines. *Fuzzy Set and Syst* 138, (2002) 271–281.
- [31] D.W. Hosmer, S. Lemeshow, R.X. Sturdivant, *Applied Logistic Regression*. 3rd Edition. Wiley (2013).
- [32] O. Kaleva, S. Seikkala, On fuzzy metric spaces, *Fuzzy Sets Syst.* 12 (1984) 215–229.
- [33] C. Kao, C.L. Chyu, Least-squares estimates in fuzzy regression analysis. *Eur J Oper Res* 148, (2003) 426–435.
- [34] I. Kramosil, J. Michálek, Fuzzy metrics and statistical metric spaces, *Kybernetika* 11 (1975) 336–344.
- [35] V. Krätschmer, Fuzzy random variables-I. Definitions and theorems, *Inform. Sci.* 15 (1978) 1–29.
- [36] V. Krätschmer, Limit distribution of least squares estimators in linear regression models with vague concepts, *J Multivariate Anal* 97, (2006) 1044–1069.
- [37] B. Kim, R.R. Bishu, Evaluation of fuzzy linear regression models by comparing membership functions, *Fuzzy Sets and Syst* 100, (1998) 343–352.
- [38] H.S. Lee, An optimal aggregation method for fuzzy opinions of group decision, *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics* 3 (1999) 314–319.

-
- [39] Y. Li, G.R. Arce, A Maximum Likelihood Approach to Least Absolute Deviation Regression, *EURASIP Journal on Applied Signal Processing* 12, (2004) 1762–1769.
- [40] M. Mizumoto, J. Tanaka, Some properties of fuzzy numbers, in: M.M. Gupta et al. (Eds.), *Advances in Fuzzy Set Theory and Applications* 153–164, North-Holland, NewYork, 1979.
- [41] M. Modarres, E. Nasrabadi, M.M. Nasrabadi, Fuzzy linear regression models with least square errors, *Appl. Math. Comput.* 163, (2005) 977–989.
- [42] M.M. Nasrabadi, E. Nasrabadi, A mathematical-programming approach to fuzzy linear regression analysis, *Appl Math Comput.* 155, 2006 873–881.
- [43] W. Näther, Regression with fuzzy random data, *Comp Stat Data Anal* 51, (2006) 235–252.
- [44] W. Näther, Regression with fuzzy random data, *Comp Stat Data Anal* 51, (2006) 235–252.
- [45] A.M. Nejad, M. Mashinchi, Ranking fuzzy numbers based on the areas on the left and the right sides of fuzzy number, *Computers and Mathematics with Applications* 61 (2011) 431–442.
- [46] D. Peidro, P. Vasant, Transportation planning with modified s-curve membership functions using an interactive fuzzy multi-objective approach, *Applied Soft Computing* 11 (2011) 2656–2663.
- [47] R.F. Phillips, Least absolute deviations estimation via the EM algorithm, *Statistics and Computing* 12, (2002) 281–285.
- [48] M.L. Puri, D.A. Ralescu, Fuzzy random variables, *J. Math. Anal. Appl.* 114 (1986) 409–422.

-
- [49] S.E. Rodabaugh, Categorical Foundations of Variable-Basis Fuzzy Topology, *Mathematics of Fuzzy Sets: Logic, Topology and Measure Theory* (U. Höhle and S.E. Rodabaugh, eds.), The Handbooks of Fuzzy Sets Series, vol. 3, Dordrecht: Kluwer Academic Publishers, 1999, pp. 273–388.
- [50] S.E. Rodabaugh, Fuzzy Real Lines and Dual Real Lines as Poslat Topological, Uniform, and Metric Ordered Semirings with Unity, *Mathematics of Fuzzy Sets: Logic, Topology and Measure Theory* (U. Höhle and S. E. Rodabaugh, eds.), The Handbooks of Fuzzy Sets Series, vol. 3, Dordrecht: Kluwer Academic Publishers, 1999, pp. 607–631.
- [51] A. Roldán, J. Martínez-Moreno, C. Roldán, On interrelationships between fuzzy metric structures, *Iran. J. Fuzzy Syst.* 10 (2) (2013) 133–150.
- [52] A. Roldán, C. Roldán, J. Martínez-Moreno, Some applications of the study of the image of a fuzzy number: Countable fuzzy numbers, operations, regression and a specificity-type ordering, *Fuzzy Sets and Systems*, D.O.I.: 10.1016/j.fss.2013.07.003
- [53] C. Roldán, A. Roldán, J. Martínez-Moreno A fuzzy regression model based on distances and random variables with crisp input and fuzzy output data: A case study in biomass production. *Soft Computing*, 16 (2012) 785–795.
- [54] A.F. Roldán-López-de-Hierro, M. de la Sen, J. Martínez-Moreno, C. Roldán-López-de-Hierro, An approach version of fuzzy metric spaces including an ad hoc fixed point theorem, *Fixed Point Theory Appl.* 2015, Article ID 2015:33, 23 páginas.
- [55] A.F. Roldán-López-de-Hierro, C. Roldán-López-de-Hierro, J. Martínez-Moreno, C. Aguilar-Peña. A family of fuzzy distance measures of fuzzy numbers. *Soft Computing*. D.O.I.: 10.1007/s00500-014-1497-0.
- [56] B. Schweizer, A. Sklar, *Probabilistic Metric Spaces*, New York, Dover Publications, 2005.

-
- [57] L. Stefanini, A generalization of Hukuhara difference and division for interval and fuzzy arithmetic, *Fuzzy Sets Syst.* 161 (2010) 1564–1584.
- [58] S.M. Taheri, M. Kelkinnama, Fuzzy linear regression based on least absolute deviations, *Iran. J. Fuzzy Syst.* 9 (1) (2012) 121–140.
- [59] H. Tanaka, I. Hayashi, J. Watada, Linear regression analysis with fuzzy model. *IEEE Trans Syst Man Cybern, SMC-12*, (1982) 903–907.
- [60] H. Tanaka, I. Hayashi, J. Watada, Possibilistic linear regression analysis for fuzzy data. *Eur J Oper Res* 40, (1989) 389–396.
- [61] L. Tran, L. Duckstein, Comparison of fuzzy numbers using a fuzzy distance measure, *Fuzzy Sets and Systems* 130 (2002) 331–341.
- [62] P. Vasant, N. Barsoum, Hybrid pattern search and simulated annealing for fuzzy production planning problems, *Comput. Math. Appl.* 60 (4) (2010) 1058–1067.
- [63] P. Vasant, *Meta-Heuristics Optimization Algorithms in Engineering, Business, Economics, and Finance* (2013).
- [64] P. Vasant, T. Ganesan, I. Elamvazuthi, Hybrid tabu search Hopfield recurrent ANN fuzzy technique to the production planning problems: a case study of crude oil in refinery industry. *International Journal of Manufacturing, Materials, and Mechanical Engineering* 2 (1) (2012) 47–65.
- [65] J. Vicente Riera, J. Torrens, Aggregation of subjective evaluations based on discrete fuzzy numbers, *Fuzzy Sets and Systems* 191 (2012) 21–40.
- [66] W. Voxman, Canonical representations of discrete fuzzy numbers, *Fuzzy Sets and Systems* 118 (2001) 457–466.
- [67] H.C. Wu, Linear regression analysis for fuzzy input and output data using the extension principle, *Comput. Math. Appl.* 45 (203) 1849–1859.

- [68] C.X. Wu, M. Ma, *The Basic of Fuzzy Analysis*, National Defence Industry Press, Beijing, 1991.
- [69] B. Wu, N.F. Tseng, A new approach to fuzzy regression models with application to business cycle analysis. *Fuzzy Sets and Syst* 130, (2002) 33–42.
- [70] L.A. Zadeh Fuzzy sets. *Inform Control* 8, (1965) 338–353.